

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 13:36:11 ; Search time 24 Seconds

(without alignments)
445.871 Million cell updates/sec

Title: US-09-813-453a-2

Perfect score: 1335

Sequence: 1 LLVIVDGNNTVGLGVYHDG.....PFLTKGLLEIYERNRGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Swisprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1114	83.4	233	1	YACB_BACSU
2	129.5	9.7	267	1	BAF_BORPE
3	90.5	6.8	4451	1	GRSB_BACBR
4	89.5	6.7	357	1	BUK_THERN
5	88.8	6.6	256	1	H161_PSEAE
6	87.5	6.6	424	1	SAHR_MERKA
7	87.5	6.5	336	1	PYRD_SALTY
8	86.5	6.5	243	1	Y004_MERJA
9	85	6.4	336	1	DMSR_ECOLI
10	83.5	6.3	204	1	YHCT_HAEIN
11	83	6.2	300	1	GSA_CLOPE
12	83	6.2	425	1	CH60_TRIVA
13	83	6.2	470	1	GALC_METJA
14	82	6.1	305	1	POLG_HCVB
15	82	6.1	3898	1	OP21_HAEIN
16	81	6.1	359	1	PRP2_YEAST
17	81	6.1	1224	1	P1MT_YERPE
18	80.5	6.0	208	1	EBGR_ECOLI
19	80.5	6.0	327	1	LUXA_VIBFI
20	80.5	6.0	410	1	RAAZ_CHURE
21	80.5	6.0	449	1	MURD_STVAM
22	80	6.0	449	1	CCP_BACSU
23	79.5	5.9	304	1	HENB_ADUAE
24	79	5.9	314	1	Y198_METJA
25	79	5.9	361	1	OP21_HAEIN
26	79	5.9	923	1	NRPI_MOUSE
27	79	5.9	931	1	SCA4_RICAU
28	78.5	5.9	337	1	TBR2_GEOCN
29	78.5	5.9	337	1	TBR2_GEOCN
30	78.5	5.9	505	1	ALMT_SCHPO
31	78.5	5.9	505	1	P1MT_SALTY
32	78	5.8	208	1	LEU2_AGRIS
33	78	5.8	469	1	LEU2_AGRIS

34	78	5.8	1032	1	ALA4_MOUSE	09w27 mus musculu
35	78	5.8	1067	1	CARB_CLOPE	08xh3 clostridium
36	78	5.8	1264	1	MOLR_ECOLI	p3345 escherichia
37	78	5.8	1342	1	RPOB_BUCAI	p5716 duchinea ap
38	78	5.8	3898	1	POLG_HCVB	p19712 hog cholera
39	77.5	5.8	204	1	YNEG_ECOLI	p77313 escherichia
40	77.5	5.8	291	1	PYRD_SULAC	008358 sulfolobus
41	77.5	5.8	352	1	EGSA_HALNI	09h49 halobacteri
42	77.5	5.8	355	1	BUK_LISTN	092h5 listeria in
43	77.5	5.8	393	1	AAT1_BACSU	p53001 bacillus su
44	77.5	5.8	428	1	PURA_CLOAB	097h87 clostridium
45	77.5	5.8	484	1	HXK1_SCHPO	009756 schizosacch

ALIGNMENTS

RESULT 1

YACB_BACSU STANDARD; PRT; 233 AA.

AC P37564;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein yacb.

GN YACB

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RC MEDLINE=96051385; PubMed=7584024;

RT Ogasawara N., Nakai S., Yoshikawa H.;

"Systematic sequencing of the 180 kilobase region of the Bacillus

subtilis chromosome containing the replication origin.";

RL DNA Res. 1:1-14(1994).

RM [2]

RM SEQUENCE FROM N.A.

RC STRAIN=168;

RC MEDLINE=98044033; PubMed=9384377;

RT Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.;

Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S.;

Boriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.;

Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.;

Choi S.K., Codani J.J., Comerford I.F., Cummings N.J., Daniel R.A.;

Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.;

Ehrlich K.D., Erlington J., Fabret C., Ferrati E., Foulger D.;

Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.;

Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.;

Guseppl G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.;

Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.;

Joris B., Karamata D., Kasahara Y., Kleber-Blanchard M., Klein C.;

Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.;

Kurita K., Lapidis A., Lardinois S., Lauber J., Lazarevic V.;

Lee S.M., Levine A., Liu H., Masuda S., Meisel C., Medigue C.;

Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nodack M.;

Noone D., O'Reilly M., Ogawa K., Ogata A., Oudega B., Park S.H.;

Parro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.;

Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.;

Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.;

Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.;

Sekiguchi J., Sekowska A., Serron S.J., Serron P., Shin B.S., Soldo B.;

Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.;

Takemaru K., Tamakoshi A., Tanaka T., Terpiltra P., Tognoni A.;

Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaret A.;

Viati A., Wambot R., Wedler E., Wedler H., Wellzenegger T.;

Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.;

Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;

"The complete genome sequence of the Gram-positive bacterium Bacillus

subtilis".

Nature 390:249-256(1997).

-1- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

RP SEQUENCE OF 1-143 FROM N.A.
 RC STRAIN-ATCC 9999;
 RX MEDLINE-90008776; Pubmed-2477357;
 RA Kretzschmar J., Krause M., Marahiel M.A.;
 RT "Gramicidin S biosynthesis operon containing the structural genes
 gtsA and gtsB has an open reading frame encoding a protein homologous
 to fatty acid thioesterases.";
 RT J. Bacteriol. 171:5422-5429(1989).
 RN [4]
 RP SEQUENCE OF 1-15, AND CHARACTERIZATION.
 RC STRAIN-Nagano;
 RX MEDLINE-92011463; Pubmed-1917901;
 RA Kurotsu T., Hori K., Kanda M., Saito Y.;
 RT Characterization and location of the L-proline activating fragment
 from the multifunctional gramicidin S synthetase 2.";
 RL J. Biochem. 109:763-769(1991).
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS PRO, VAL, ORN AND LEU.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
 CC ANTIOTIC GRAMICIDIN S (D-PHE-PRO-VAL-ORN-LEU)².
 CC -1- SUBUNIT: LARGE MULTIMERIC COMPLEX OF GRS AND GRSB.
 CC -1- DOMAIN: CONSISTS OF FOUR MODULES, AND HARBOURS A PUTATIVE
 CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
 CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
 CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
 CC ADENYLATION, THIOATION, CONDENSATION (NOT FOR THE INITIATION
 CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
 CC (OPTIONAL).
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X61658; CAA53838.1; -;
 CC EMBL: M29703; AA58719.1; -;
 CC EMBL: X15577; CAA33604.1; -;
 CC EMBL: D00938; BAA00778.1; -;
 CC PIR: S20542; YGBCS2.
 CC HSSP: P14687; 1AMU.
 CC InterPro: IPR000873; AMP-bind.
 CC InterPro: IPR001242; Condensatn.
 CC InterPro: IPR003880; Pantine-attach.
 CC InterPro: IPR003379; Ser-estr-act.
 CC InterPro: IPR001031; Thioesterase.
 CC Pfam: PF00501; AMP-binding; 4.
 CC Pfam: PF00550; PP-binding; 4.
 CC Pfam: PF00668; Condensation; 4.
 CC Pfam: PF00975; Thioesterase; 1.
 CC PRINTS: PR00154; AMPBINDING.
 CC PROSITE: PS00012; PHOSPHOPANTETHEINE; 4.
 CC PROSITE: PS00455; AMP-BINDING; 4.
 CC PROSITE: PS0075; ACP_DOMAIN; 4.
 CC Lgase; Antibiotic biosynthesis; Multifunctional enzyme;
 CC Repeat; Phosphopantetheine; Hydrolase.
 CC KW
 CC INIT MET 0
 CC REPEAT 466 1043 DOMAIN 1 (PROLINE-ACTIVATING).
 CC REPEAT 1521 2080 DOMAIN 2 (VALINE-ACTIVATING).
 CC REPEAT 2538 3135 DOMAIN 3 (ORNITHINE-ACTIVATING).
 CC REPEAT 3591 4173 DOMAIN 4 (LEUCINE-ACTIVATING).
 CC DOMAIN 975 1042 ACYL CARRIER (ACP) 1.
 CC DOMAIN 2011 2078 ACYL CARRIER (ACP) 2.
 CC DOMAIN 3057 3124 ACYL CARRIER (ACP) 3.
 CC DOMAIN 4095 4162 ACYL CARRIER (ACP) 4.
 CC BINDING 1005 1005 PHOSPHOPANTETHEINE (BY SIMILARITY).

FT BINDING 2041 2041 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 3087 3087 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT BINDING 4125 4125 PHOSPHOPANTETHEINE (BY SIMILARITY).
 FT CONFLICT 274 274 H -> D (IN REF. 2).
 FT CONFLICT 418 418 K -> R (IN REF. 2).
 FT CONFLICT 654 664 NAVLYCTTKR -> TCSFDVQOEI (IN REF. 2).
 FT CONFLICT 941 946 HVRLHL -> OLPLTP (IN REF. 2).
 SO SEQUENCE 4451 AA; 510036 MW; E0029C9B51F5A4B7 CRC64;
 Query Match 6.88; Score 90.5; DB 1; Length 4451;
 Best Local Similarity 22.78; Pred. No. 44;
 Matches 56; Conservative 35; Mismatches 77; Indels 79; Gaps 14;
 QY 70 MEALERMCTKYFHIEPOLVCGMKTGILKIKDNPEVGS-----ADRIY 112
 Db 1396 MFVLQNDNRKSEFEYEQITITTYVRSRNSKFDLLEVSSEONEILLCEYTKLFTDITY 1455
 QY 113 NAVAA-----IH-LYGNPLIV-----DFG-TATTCYIDENQOYM-- 146
 Db 1456 ERMAGHFLQILHAIVGNPTIITSEIEILSEBEKHILFEFNDTKTYPLCKOFKDYLNRR 1515
 QY 147 -----GGAIAFG---ITISTELYSRAALPRI---EITRPNIICKNTVSAHQGILFCY 196
 Db 1516 WRRADHVAVGKDOTLYRELNERANQAVLQKGVQPNVIG--LLVERSPMLVG- 1572
 QY 197 VGOVEGIYK-----RMKWAQKDLKV---ITGGLAPLANESDCIDIDY 238
 Db 1573 ---IMGILKAGAVLPDPEYPADRIISYMT-QDCGVRLMLQOHLISLVHDFECV-ILD 1627
 QY 239 PFLTLKG 245
 Db 1628 EDSLKYG 1634
 RESULT 4
 BUK_THETN
 ID BUK_THETN STANDARD; PRT; 357 AA.
 AC Q8R832;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable butyrate kinase (EC 2.7.2.7) (BK) (Branched-chain carboxyl)c
 DE acid kinase).
 GN BUK OR TTE2201.
 OS Thermoaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermoaerobacteriales;
 OC Thermoaerobacteriaceae; Thermoaerobacter.
 OX NCBI_Taxid=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MB4 / JCM 11007;
 RX MEDLINE-21992816; Pubmed-11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome.";
 RL Genome Res. 12:689-700(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + 2-butanone - ADP + butanoyl phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE ACETOKINASE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF013165; AAM25356.1; -;
 CC DR PROSITE: PS01075; ACETATE_KINASE_1; 1.
 CC DR PROSITE: PS01076; ACETATE_KINASE_2; 1.
 CC KW
 CC Transferase; Kinase; Complete proteome.

Seq	Sequence	357 AA	39217 MW	05CDDE09A2E7EF3F CXC64
QY	Query Match	6.7%	Score 89.5	DB 1; Length 357;
DB	Best Local Similarity	19.9%	Pred. No. 2.8;	
Matches	66; Conservative	43;	Mismatches	94; Indels 129; Gaps 16;
QY	1	LLIVDVCNTNTNVGLGVYHDGKLEFYWMRLTETSHTKE-----DEGMIILRSFD---	48	
DB	5	LILVIVNPSTSTSKVAVFDEKPEVF--RETLNHSSTELSKYKRRIIDQFEFTQATILDMLK	61	
QY	49	HSGLMFEQIDIGI-----ISS-----	64	
DB	62	EKGISLSQIDAIYVGGGLPKIEGSTYIVNEMKMLDCKAERGHASNLGAIATLAKKE	121	
QY	65	-----VPPIMFLERMC-----TKYH-IEPQIVPGMKGLNITKYNPKEV	106	
DB	122	HNIPAYIVDPVAVDELEVARITGLPEIEKOSIFALNOKALARRIADSLGRKDE---	177	
QY	107	GADRIYVNAVAIHLYGNPLIVDFSTATTQCYIDENKQYMG-GARAPGTTISTEALYSRA	165	
DB	178	-----VNLIIA-HLGGG-----ISVAHKKGRIVYDNDALNGGPRSP-----ERA	217	
QY	166	AKLPRIETRPDNIIGKNTVSAMQ-----SGILFSYVG-----QVEGIV	204	
DB	218	GLPLVLDLVKLT-CYSGKTYFEEMKKLLIGKGVIAHLTNDVREYVKMIENGDKNAEILT	276	
QY	205	KRMKQAKQADCKVIATVGTGLAPLANESPCIDI	236	
DB	277	DAMAQTAKER-----GSMAYVLKKQYVALGI	303	
RESULT 5				
H161_PSEAE				
ID	H161_PSEAE	STANDARD:	PRT:	256 AA.
AC	O9HU44:			
AD	15-JUN-2002	(Rel. 41, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Imidazole glycerol phosphate synthase subunit hisF1 (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF1) (imgp synthase subunit hisF1) (IGPS subunit hisF1).			
GN	HISF1 OR PA5140.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=287;			
RA	[1]			
RP	SEQUENCE FROM N.A.			
STRAIN	ATCC 15692 / PA01;			
RA	MEDLINE=204373737. PubMed=10984043;			
RA	Stover C.K., Pham X.-O.T., Ervin A.L., Mizoguchi S.D., Warriner P.,			
RA	Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.U., Lagrou M.,			
RA	Gader R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A.A., Iarbig K., Lam R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;			
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an			
RT	opportunistic pathogen."			
RL	Nature 406:959-964(2000).			
CC	-1- FUNCTION: IGS catalyzes the conversion of PEPAR and glutamine to			
CC	IGP, AICAR and glutamate. The hisF subunit catalyzes the			
CC	cyclization activity that produces IGP and AICAR from PEPAR using			
CC	the ammonia provided by the hisH subunit (By similarity).			
CC	-1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-			
CC	ylamino)methylideneamino]-1-[(5-phosphoribosyl)imidazole-4-			
CC	carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-			
CC	aminimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.			
CC	-1- PATHWAY: Histidine biosynthesis; fifth step.			
CC	-1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC DRC EMBL, AE004927; AAC08525.1.
CC DR InterPro; IPR003009; FMN_enzyme.
CC DR InterPro; IPR000570; His_biosynth.
CC DR Pfam; PF00977; His_biosynth.1.
CC DR TIGRFAMs; TIGR00735; hisF.1.
CC Hisidine biosynthesis; Lyase; Complete proteome.
CC ACT_SITE 12
CC ACT_SITE 12 POTENTIAL.
CC ACT_SITE 131
CC ACT_SITE 131 POTENTIAL.
CC SQ SEQUENCE 256 AA; 27131 MW; 64866A/6CD308FAA7 CnC64;
CC -----
Query Match 6.6%; Score 88; DB 1; Length 256;
Best Local Similarity 22.7%; Pred. No. 2.5;
Matches 60; Conservative 45; Mismatches 79; Indels 80; Gaps 15;
CC
QY 5 IDVGNTNYLYGYHGGKLEHYHRIETSRKRTDEDEGMILRLSFHDSGLMFEQIDDIITSS 64
DB 11 IDVNGRVVYKGYKPFENIRAGDPVEIARR-----YDQGA--DETFLDDTA 55
QY 65 VV---PIWFALEPRCKTKEFHIEPQIVGSGMKTGINRYNKEKEGADRIYAAVAHLY 121
DB 56 SYDGRDITLHYVERMASQVF--IPLTVGGVAYSVDIR--NLINAGADKVSINTAAVF-- 109
QY 122 GNPPLYVD---FGATITCYIDEKKQYMGALAGCITITSTALXSRAPKPRIET---- 173
DB 110 -NPEFVEGEADRFSGQCIYVAIDARK-----VSAQG-----EAPRWEIFTHG 150
QY 174 ----TRPDNII-----GKNTVSAM-QSGIIFGY-VGVGEGIVKRMKQAKQDKLV 217
DB 151 GKRPGLDAVLAAMAKMEDGAGEIILLTSMDDQGVASGVDLGVTRAISEAV-----NNPV 204
QY 218 IATGGL-----APLIANESDPT 234
DB 205 IASGVGNLHPLAAGILLEKADAV 228
CC -----
RESULT 6
SAHH_METKA
ID SAHH_METKA STANDARD: PRT; 424, AA.
AC P58855;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenosylhomocysteinase [EC 3.3.1.1] (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
DE GN AHCY OR MK0368.
OS Methanopyrus kandleri.
OS Archaea; Euryarchaeota; Methanopyrus; Methanopyrales; Methanopyraceae;
OS Methanopyrus.
OC NCBI_TaxID=2320;
CC
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina O.V., Shaknova V.V., Belova G.I., Arvind L.,
RA Natarle D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malyskh A.G., Koonin E.V., Kozayvkin S.A.; Methanopyrus kandleri AV19
RT "The complete genome of hyperthermophilic Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -I- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC + L-homocysteine.
CC -I- COFACTOR: NAD (By similarity).
CC -I- PATHWAY: Activated methyl cycle.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.

```

```

RT LT2.":
RL Nature 413:852-856(2001).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroxycitrate + O(2) = orotate +
CC H(2)O(2).
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- PATHWAY: Pyrimidine biosynthesis: fourth step.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DIHYDROXYOTATE DEHYDROGENASE FAMILY.
CC SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55636; CA39161.1; -
DR EMBL: AE008746; AAL19991.1; -.
DR PIR: S13824; S13824.
DR ScyGene: SG10326; PYD.
DR InterPro: IPR001295; DHO_dh.
DR Pfam: PF01180; DHOdebase; 1.
DR TIGRFAMS: TIGR01036; pyrd_sub2; 1.
DR PROSITE: PS00911; DHODEHASE_1; 1.
DR PROSITE: PS00912; DHODEHASE_2; 1.
DR Pyrimidine biosynthesis: Oxidoreductase; Flavoprotein; FMN;
KW Complete proteome.
FT NP_BIND 290
FT FT 298 FMN (POTENTIAL).
SQ SEQUENCE 336 AA; 36739 MM; 783F59D28DE6398 CRC64;

Query Match 6.5%; Score 87; DB 1; Length 336;
Best Local Similarity 21.6%; Pred. No. 4.2;
Matches 58; Conservative 36; Mismatches 79; Indels 96; Gaps 13;

QY 37 DEFGILRLSDPHSLT-MPEOI-----DGIILSYVPPIFALERMCTKTFHIEPOLVG 89
DB 105 DAEGILNMGRRNNGVDLVENKKAHPDGIL----- 136
QY 90 PGMKTGLNI-KYDNPKEVGADRIYNAVAALHLYGNPLIV-----VDFGPAAT 135
DB 137 -----GINIGKNKPTPYENGKDYLICMEKYAVAGYIAINISSPNPGLTLQYGA-- 189
QY 136 YCYIDE-----NNO-----YMGALAPGTTISTEALYSRAAKPRIETRPDNI 180
DB 190 ---LDDLLTAIRKNQNDLQAIHMKYVPAVAATADLCEELIQVADSLLRHNI--DGYI 243
QY 181 GKNTV-----SAMQSIGLFGYGVGVG--IVKRMKQAKQADKLATVATGGLAPLA 228
DB 244 ATNTLTDRSLVQGMKNCQQTGSLGRPLQGLKSTELIRRLSDELKQLPLTIGVGSDISVA 303
QY 229 NE-----SCDIDVDFPLTKGLLEIYE 251
DB 304 AREKIAAGATLVQITSGFI-FKGPPLIKE 331

RESULT 8
Y004.METJA
ID Y004.METJA STANDARD; PRT; 243 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0004.
GN MJ0004.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
NCBI_TaxID=2190;
[1]

```

SEQUENCE FROM N.A.
 RP STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE-96337999; PubMed-8688087;
 RA Bult C.J., White O., Olsen G.O., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodok A.,
 RA Scott J.L., Geophagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1 SIMILARITY: A.FERMENTANS ACTIVATOR OF (R)-HYDROXYGLUTARYL-COA
 DEHYDRATASE (HGDC), TO E.COLI XJLI AND M.JANNASCHII MJ0800.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U67459; AAB97965.1; -
 DR HSSP: P11568; 1H0X.
 DR TIGR: MJ0004;
 DR InterPro: IPR002731; ATPase_BadP.
 DR Pfam: PF01869; BcrAD_BadP.1.
 DR ProDom: PD006344; ATPase_BadP.1.
 DR TIGRfams: TIGR00241; CoA_E_activ.1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 243 AA; 26897 MW; 288C4F4190575E53 CRC64;
 SQ
 Query Match 6.5%; Score 86.5; DB 1; Length 243;
 Best Local Similarity 21.3%; Pred. No. 3.2;
 Matches 54; Conservative 44; Mismatches 81; Indels 75; Gaps 12;
 QY 1 LLAIVDGNNTVLYGVYHDKLEKLEWHRIETSRHKTDEDFEGLMRLSLFDSHGLMEQIDGI 60
 DB 1 MILDIDVSTTKVMLEDSKLIW-----YKIED-----IGV 32
 QY 61 IISVVPIMALERMCTKXPHIEPOIVGPMKGLIKYDKNPEVGAADRIYNAVAI-- 118
 DB 33 VIEE---DILKWKKEIEQKIPD-KIVATG-----YGRKVSADKIVPEYVIALGR 80
 QY 119 ---HLYGNPLIVDF-GTATTCYIDENKQYMGGAIAPIGIIISTEALYSRA--AKLPRI 171
 DB 81 GANFFNADGVIDIGODTFVKLIDKNKGVDFILSDKCAAGTKFLEKALDILKIDKN 140
 QY 172 EIT--RPDNIIGKNTVSM-----QSGILRG---YGYGVGYIKRKMKWQ 210
 DB 141 EINKYKSDNIKISSMCVFAESEITISLKRKVEGILMGVSEITINRVIPMTNRKLIQ 200
 QY 211 AKODLKVIATGGLA 224
 DB 201 -----NIVFSGGVA 209
 RESULT 9
 PYRD_ECOLI STANDARD: PRT: 336 AA.
 AC P05021;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dihydrocrotonate dehydrogenase (EC 1.3.3.1) (Dihydrocrotonate oxidase)
 DE (DHODEHASE) (DHODASE) (DHOD)
 GN PYRD OR B0945 OR Z1294 OR ECS1029.
 OS Escherichia coli, and
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-7.
 RX MEDLINE-85285014; PubMed-2992959;
 RA Larsen N.J., Jensen K.F.;
 RT "Nucleotide sequence of the pyrd gene of Escherichia coli and
 RT characterization of the flavoprotein dihydrocrotonate dehydrogenase.";
 RL Eur. J. Biochem. 151:59-65(1985).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97061202; PubMed-8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajinara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinouchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE-21074935; PubMed-11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatter F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0157:H7 / RIMD 0509952;
 RX MEDLINE-21156231; PubMed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- CATALYTIC ACTIVITY: (S)-dihydrocrotonate + O(2) = crotonate +
 CC H(2)O(2).
 CC -1- COFACTOR: FMN.
 CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE DIHYDROCRONATE DEHYDROGENASE FAMILY.
 CC SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X02826; CAA26594.1; -
 DR

DR EMBL: AE000196; AAC74031.1; -
 DR EMBL: D90732; BAA35700.1; -
 DR EMBL: AF005284; AAC55431.1; -
 DR EMBL: AF002553; BAB34432.1; -
 DR PIR: A23109; DEECDO.
 DR SWISS-2DPAGE: P05021; COLI.
 DR Ecocore: EG10807; PYRD.
 DR InterPro: IPR001295; DHO.dh.
 DR InterPro: IPR003009; FMN_enzyme.
 DR Pfam: PF01180; DHODHase.1.
 DR TIGRfam: TIGR01036; PYRD_sub2.1.
 DR PROSITE: PS00911; DHODHASE.1; 1.
 DR PROSITE: PS00912; DHODHASE.2; 1.
 DR Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
 KW Complete proteome.
 FT NR_BIND 290 298 FMN (POTENTIAL).
 FT SEQUENCE 336 AA; 36774 MW; 973227EAB6B83622 CRC64;
 Query Match 6.48; Score 85; DB 1; Length 336;
 Best Local Similarity 25.08; Pred. No. 6.2;
 Matches 51; Conservative 27; Mismatches 70; Indels 56; Gaps 11;
 QY 95 GLNI--KYDNKFWGADRIYNAVAHLYGNPLIV-----VDECTATTCYID 140
 DB 137 GINIGKMKDTPVEQKDDYLICMEKIYAYAGYIAINISPTPLRLTLYGEA-----LD 191
 QY 141 E-----NKO-----YMGALAPCITITSTALYSRAKLPRIETTRDNILGNT- 184
 DB 192 DLTAINKKNDLQAMHMKIVPIAVKTAIPDLSEELQVADSLVRHNT---DGIATNTT 248
 QY 185 ----VSAM-----OSGILFYGVGVEG--IVRKMKQAKODLKVIATGGLAPLANE--- 230
 DB 249 LDRSLVCGMKMKDQGTGSLGRPLDLKSTELIRLRLSLNGLRLPLIGVGIDSVYANAEKI 308
 QY 231 ---SDCDIDYDPFLTLGLGLIYE 251
 DB 309 AAGASLVQIVSGFI-FKGPPLIKE 331
 RESULT 10
 DMSB_ECOLI STANDARD; PRT; 204 AA.
 AC P18776; F77745;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Anaerobic dimethyl sulfoxide reductase chain B (DMSO reductase iron-
 sulfur subunit).
 OS DMSB OR B0895.
 NC Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID:562;
 RX NCBI [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-9.
 RC STRAIN-K12 / CG00;
 RC MEDLINE-97426617; PubMed-3062312;
 RA Blious P.T., Cole S.T., Anderson W.F., Weiner J.H.;
 RA "Nucleotide sequence of the dmsABC operon encoding the anaerobic
 RT dimethylsulphoxide reductase of Escherichia coli.";
 RL MOL. Microbiol. 2:785-795(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RC MEDLINE-97426617; PubMed-9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN-K12;
 RX MEDLINE-97061202; PubMed-8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12,7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP EPR SPECTROSCOPY OF IRON-SULFUR CLUSTERS.
 RX MEDLINE-91070067; PubMed-2174699;
 RA Cammack R., Weiner J.H.;
 RT "Electron paramagnetic resonance spectroscopic characterization of
 RT dimethyl sulfoxide reductase of Escherichia coli.";
 RL Biochemistry 29:8410-8416(1990).
 RN [5]
 RP MUTAGENESIS.
 RX MEDLINE-91355180; PubMed-1653010;
 RA Rothery R.A., Weiner J.H.;
 RT "Alteration of the iron-sulfur cluster composition of Escherichia
 RT coli dimethyl sulfoxide reductase by site-directed mutagenesis.";
 RL Biochemistry 30:8286-8305(1991).
 CC -1- FUNCTION: Electron transfer subunit of the terminal reductase
 CC during anaerobic growth on various sulfoxide and N-oxide
 CC compounds.
 CC -1- COFACTOR: BINDS 4 4FE-4S CLUSTERS.
 CC -1- SUBUNIT: THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA, THE
 CC REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSG, A
 CC MEMBRANE ANCHOR PROTEIN.
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
 CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 DR EMBL: J03412; AAA83844.1; -
 DR EMBL: AE000191; AAC73981.1; -
 DR EMBL: D90727; BAA35627.1; -
 DR PIR: S03786; S03786.
 DR HSSP: P00195; 1CLF.
 DR Ecocore: EG10233; dmsB.
 DR InterPro: IPR001450; 4Fe4S_ferredoxin.
 DR Pfam: PF00037; fer4; 1. 4Fe4S_ferredoxin.
 DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 1.
 KW Electron transport; 4Fe-4S; Iron-sulfur; Complete proteome.
 FT INIT_MET 0 0
 FT METAL 13 13 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 16 16 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 19 19 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 22 22 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 23 23 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 66 66 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 69 69 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 74 74 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 78 78 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 98 98 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 101 101 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 104 104 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 108 108 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 125 125 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 128 128 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 140 140 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 144 144 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 144 144 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT MUTAGEN 101 101
 FT CONFLICT 169 169
 P -> PRA (IN REF. 1).
 FROM MENADINOL TO DMSO.

HSP: P24630; 2GSA.
 DR InterPro: IPR000954; AminoLtrn_3.
 DR InterPro: IPR004639; HemL.
 DR Pfam: PF00202; aminoLtrn_3; 1.
 DR TIGRfams: TIGR00713; hemL; 1.
 DR PROSITE: PS00600; AA_TRANSFER_CLASS_3; 1.
 DR Porphylin biosynthesis; Isomerase; Pyridoxal phosphate;
 KM Complete proteome.
 FT BINDING 265 265 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 FT CONFLICT 73 73 E -> K (IN REF. 1).
 FT CONFLICT 323 323 N -> D (IN REF. 1).
 FT CONFLICT 356 356 I -> M (IN REF. 1).
 FT CONFLICT 422 422 T -> S (IN REF. 1).
 SQ SEQUENCE 425 AA; 46881 MW; B23DF062CE39B9E2 CRC64;

Query Match 6.2%; Score 83; DB 1; Length 425;
 Best Local Similarity 22.6%; Pred. No. 12;
 Matches 43; Conservative 21; Mismatches 52; Indels 74; Gaps 10;

11 NTVLGVYHDKLEYHRIETSRHKTDEFGMLRLSLFDHSGLMFEQIDGIISVV-----66
 175 NTLLIGLYNDEKO-----VEELFEKYG---NDIAGIIIEPVAGNMG 211
 67 -----PIPFALERMCTKY-----FHIEP-----QIVGPGM 92
 212 VVKCDKFKRKLRELCDDKCALLIFEVWCGFRVAYKGAQTLEDPKPDLYTAKINGGGL 271
 93 KKGILNKYKNEKREV--GARIYNAVAHILKGNPLVDEGFTATTCYIDENKQY-----145
 272 PGC---AYGGRREIMENLSPLGGVYQAGTSGNP-IYMSAGATVAK-KLYENPSYNNHIE 326
 146 -MGAIAPCI 154
 327 KTSKLEKCV 336

RESULT 13
 CH60_TRIVA STANDARD; PRT; 470 AA.

AC 095058;
 DT 15-JUL-1998 (rel. 36, Created)
 DT 15-JUL-1998 (rel. 36, Last sequence update)
 DT 15-JUL-1998 (rel. 36, Last annotation update)
 DE Hydrogenosomal chaperonin HSP60 (Protein Cpn60) (GroEL protein) (Heat shock protein 60).
 GN HSP60.
 OS Trichomonas vaginalis.
 NC Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae;
 NC Trichomonadinae; Trichomonas.
 NCBI_TaxID=5722;

RP MEDLINE=96382521; PubMed=8790385;
 RA Bui E.T., Bradley P.J., Johnson P.J.;
 RT "A common evolutionary origin for mitochondria and hydrogenosomes";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:9651-9656(1996).
 CC -1- FUNCTION: IMPLICATED IN HYDROGENOSOMAL PROTEIN IMPORT AND
 MACROMOLECULAR ASSEMBLY. MAY FACILITATE THE CORRECT FOLDING OF
 IMPORTED PROTEINS. MAY ALSO PREVENT MISFOLDING AND PROMOTE THE
 REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED
 UNDER STRESS CONDITIONS IN THE HYDROGENOSOME (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: HYDROGENOSOMAL.
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; U26966; AAB17250.1;

HSP: P06139; 1GRU.
 DR InterPro: IPR001844; Chaperonin_Cpn60.
 DR InterPro: IPR002423; Cpn60/TCF-1.
 DR Pfam: PF00118; cpn60_TCF1; 1.
 DR PRINTS: PR00298; CHAPERONIN60.
 DR PRINTS: PR00304; TCOMPLEXTCP1.
 DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
 KM Chaperone; ATP-binding; Hydrogenosome.
 SQ SEQUENCE 470 AA; 50614 MW; 6D501801045605AF CRC64;

Query Match 6.2%; Score 83; DB 1; Length 470;
 Best Local Similarity 18.8%; Pred. No. 13;
 Matches 51; Conservative 49; Mismatches 76; Indels 96; Gaps 14;

28 IETSRH-----KTED-EGGMILRLSLFDHSGLMFEQIDGIISVVP---IPIFALERM 77
 4 IEAKHFTFAFAARLKLKGSAR---DHLLGVEKLAADVSTLSPKGRNWIPLP---57
 78 TKFHEIPQIVGPGMTGLNIRK-DNPKREVADRIYNAVAH-----L 120
 58 ----YGPVKYTKGIVAVASIEFKDKQNGQOLVINAKTNDVAGDGTTLTLTREL 113
 121 YGNPLIVDEGATTCYCYIDENKQYMGALAPGITISTEALYSRAAKLPRIETRPDNI 180
 114 YRESIKALSAG-----LDPNK-----VRKGMTLRVAVVLEKSTK-KVSSPDEIF 159
 181 GKNTVSAMOS-----GILFGYVGQ-----VSGI-----203
 160 NVATISANSEKIGLILADAFKAVNGEYIVTVMGKRFHELETVOGMKIDRGYISSYQ 219
 204 --VKRMKQWAKQ-----DLKVIATGGLPLI 227
 220 NDKSMKCEYENPIITDINKINSFAQIPL 251

RESULT 14
 GALE_METUA STANDARD; PRT; 305 AA.

AC 057664;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Putative UDP-glucose 4-epimerase (EC 5.1.3.2) (galactowaldenase) (UDP-galactose 4-epimerase).
 GN MJ0211.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 NCBI_TaxID=2150;

RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Uteback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Haurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL Science 273:1058-1073(1996).
 CC -1- CATALYTIC ACTIVITY: UDP-glucose -> UDP-galactose.
 CC -1- COFACTOR: NAD.
 CC -1- PATHWAY: Galactose metabolism; third step.
 CC -1- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U67477; AAB98196.1; -
 DR HSSP: P09147; IXL.
 DR TIGR: M0211; -
 DR InterPro: IPR001509; Epimerase_Dh.
 DR Pfam: PF01370; Epimerase; 1.
 DR Hypothetical protein: Isomerase; NAD: Galactose metabolism;
 KW Complete proteome.
 FT NP_BIND 1 32 NAD (POTENTIAL).
 SQ SEQUENCE 305 AA; 34542 MW; 2756773CF95D50BD CRC64;
 Query Match 6.1%; Score 82; DB 1; Length 305;
 Best Local Similarity 24.7%; Pred. No. 9.8; Mismatches 52; Indels 72; Gaps 12;
 Matches 49; Conservative 25; Mismatches 52; Indels 72; Gaps 12;
 QY 99 KYDNKEVGAADRIYNAVAHLYGNPLIVDFGATTCYCIDEN-----KQYMG 147
 104 KYD-----IDKIYVASSGSGAVYGE-----NLPVDENHPINPLSPYGLSKYVG 147
 QY 148 GAIAAGITISREALYSRAAKLPRIE-ITRPDNIIGK-----NTYSAMQS 190
 148 EBYI-----KLYNR---LYGLEIATLRYSNYGEGKODPKGEGAVISIFIDKMLKMQS 196
 QY 191 GILFG-----YGVGV-EGIVKRMKQAKODKVIATGGLADPLINESCDIYDPF 240
 197 PIIRBDGNOTRDFYVGVGDAKANLMALNW--KNEIVNIGTKRETSV---NELPITLHE 250
 QY 241 LTLKGLLEIYERNRVGSV 258
 DB 251 IGFRG-EAIVDKPREGEV 267
 RESULT 15
 POLG_HCVB STANDARD; PRT; 3898 AA.
 ID POLG_HCVB
 AC P21530;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Genome polyprotein.
 OS Hog cholera virus (strain Brescia) (Swine fever virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_Taxid=11098;
 RN [1]
 SEQUENCE FROM N.A.
 MEDLINE=90281581; PubMed=2162104;
 Moormann R.J.M., Warmerdam P.A.M., van der Meer B., Schaaper W.M.M.,
 Wensvoort G., Hulst M.M.;
 "Molecular cloning and mapping of the genomic region encoding envelope
 strain Brescia and mapping of the genomic region encoding envelope
 protein E1.";
 RT Virology 177:184-198(1990).
 RL Virology 177:184-198(1990).
 CC -1- FUNCTION: PESTIVIRUS P80 (P125) MAY BE A BIFUNCTIONAL PROTEIN
 CC WITH HELICASE AND PROTEASE ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: THE GP51-GP54 PROTEIN IS ANCHORED TO THE
 CC VIRAL ENVELOPE.
 CC -1- SIMILARITY: TO BOVINE VIRAL DIARRHEA VIRUS GENOME POLYPROTEIN.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S31.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M31768; AAA43843.1; -
 DR PTR: A35317; GNMVHB.

DR HSSP: P27958; 1AIV.
 DR MEROPS: C53.001; -
 DR MEROPS: S31.001; -
 DR InterPro: IPR000280; Cdvir_endptasep80.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002166; HCV_RORP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR001568; RNase_T2.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; HCV_RORP; 1.
 DR PRINTS: PR00729; CDVENDOPTASE.
 DR SMART: SM00490; HELIC_C; 1.
 DR PROSITE: PS00531; RNASE_T2_2; UNKNOWN_1.
 KW Polyprotein; Glycoprotein; Transmembrane; Hydrolase; Serine protease;
 KW Helicase.
 FT CHAIN 21 2267 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 2268 2500 GP42 (E2) (POTENTIAL).
 FT CHAIN 2501 2689 GP31 (E3) (POTENTIAL).
 FT CHAIN 2690 21060 GP51-GP54 (ENVELOPE PROTEIN E1).
 FT CHAIN 21611 22111 P80 (POTENTIAL).
 FT TRANSMEM 1032 1048 POTENTIAL.
 FT ACCT_SITE 1658 1658 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACCT_SITE 1695 1695 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1752 1752 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 805 805 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1713 1713 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2134 2134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2217 2217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2419 2419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2494 2494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2787 2787 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2815 2815 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2891 2891 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3103 3103 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3211 3211 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3316 3316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3689 3689 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3698 3698 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3794 3794 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3898 AA; 438423 MW; EC6B207A09D59FD CRC64;
 Query Match 6.1%; Score 82; DB 1; Length 3898;
 Best Local Similarity 21.2%; Pred. No. 1.9e+02; Mismatches 73; Indels 64; Gaps 11;
 Matches 46; Conservative 34; Mismatches 73; Indels 64; Gaps 11;
 QY 20 KLEIYHW-RIEHSRKTDEDEGMILR-----SLPFGSGLMFEIROID---GIISVVP 67
 2113 GSKDYHDLQAKRYGIEDGINITKSPREMANYSLYEEDSLMITQLILNNLISELP 2172
 QY 68 PIMFALERMCTRYFHIEP-----QIVGFMKTG-LINKYDNPKREVGAADRIYNA 114
 2173 ---MAVKNNIMARTDHPEDIQALVNSYENQVPLFPKINGEVTSDYVTFLLNARKLDD 2229
 QY 115 VAAHLYGNPLIVDFGATTCYCIDENKQYMGAI-A-----PGITISREALYSRAALP 169
 2230 V-----PPVYATDEDEDLAVELLGLDWDPEGNOGVVEA--GAALK- 2267
 QY 170 REITRPDNIIGKNTVSAMSGILFGYGVGVEIYKR 206
 DB 2268 -----QVGLSTAEANALLVALFVG-VYALSKR 2295

Tue Apr 22 16:35:09 2003

us-09-813-453a-2.rsp

Page 11

Search completed: April 21, 2003, 13:40:34
Job time : 28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 13:36:36 ; Search time 84 Seconds
(without alignments)
632.859 Million cell updates/sec

Title: US-09-813-453A-2
Perfect score: 1335
Sequence: 1 LLIVDVGNTVTVLGVYHDG.....PFLTKGLIELYERNRVGSV 258
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1036	77.6	258	2	09F985
2	987	73.9	254	16	09K9H5
3	852	63.8	259	16	08YAC5
4	847	63.4	259	16	092F54
5	775	58.1	253	16	097EB4
6	773	57.9	255	16	08R7M2
7	737	55.2	259	16	08XHL5
8	655	49.1	265	16	09XN6
9	577	43.2	261	16	09A6Z1
10	519.5	38.9	274	16	09CD56
11	494.5	37.0	272	16	006282
12	458	34.3	256	16	08RFE4
13	428	32.1	262	16	09R554
14	405.5	30.4	246	16	09WZY5
15	338.5	25.4	212	2	032514
16	327.5	24.5	273	16	083446

17	271.5	20.3	262	16	051477	051477 borrelia bu
18	171	12.8	295	16	08Y2M4	08Y2M4 ralsicoma s
19	170	12.7	276	16	08YOD7	08YOD7 anabena sp
20	166	12.4	56	2	P94305	P94305 bacillus ps
21	163	12.2	257	16	P74045	P74045 synechocyst
22	154.5	11.6	229	16	067753	067753 aquifex aeo
23	150	11.2	592	16	09JXF1	09JXF1 nelsaria m
24	150	11.2	592	16	09JW17	09JW17 nelsaria m
25	134.5	10.1	242	16	09PC14	09PC14 xyella fas
26	133	10.0	248	16	09HWC1	09HWC1 pseudomonas
27	117	8.8	224	16	098093	098093 mycoplasma
28	116.5	8.7	223	16	092KX6	092KX6 helicobacte
29	109.5	8.2	223	16	025533	025533 helicobacte
30	109	8.2	209	16	09P1A9	09P1A9 campylobact
31	104.5	7.8	597	17	030225	030225 archaeoglob
32	100	7.5	318	10	09LYY3	09LYY3 arabidopsis
33	95	7.1	320	16	08U6P6	08U6P6 agrobacteri
34	94.5	7.1	858	12	09Q7D8	09Q7D8 marek's dis
35	93.5	7.0	895	17	0972N0	0972N0 sulfolobus
36	93	7.0	1014	5	096276	096276 plasmodium
37	91.5	6.9	485	16	08RPP6	08RPP6 fusbacteri
38	91.5	6.9	828	16	09PBB9	09PBB9 xyella fas
39	90	6.7	676	12	091WV8	091WV8 lumby skin
40	89.5	6.7	357	16	08R832	08R832 thermoaer
41	89.5	6.7	858	12	09E115	09E115 melagrid h
42	89.5	6.7	858	12	09DPS9	09DPS9 melagrid h
43	87.5	6.6	205	16	08ZB18	08ZB18 salmonella
44	87.5	6.6	287	16	0980Y2	0980Y2 mycoplasma
45	87	6.5	336	16	08Z7S9	08Z7S9 salmonella

ALIGNMENTS

RESULT 1
ID Q9F985 PRELIMINARY; PRT; 258 AA.
AC Q9F985;
DT 01-MAR-2001 (TREMUREL. 16, Created)
DT 01-MAR-2001 (TREMUREL. 16, Last sequence update)
DT 01-JUN-2002 (TREMUREL. 21, Last annotation update)
DE Putative 32 kDa replication protein.
OS Bacillus steareophilus.
OC Bacteria; Firmicutes; Bacillus/clostridium group; Bacillales;
OC Geobacillus.
CX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V;
RA Vasquez C., Pichuanes S., Saavedra C.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF198621; AAC28531.1; -;
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 77.6%; Score 1036; DB 2; Length 258;
Best Local Similarity 77.9%; Pred. No. 3.9e-84;
Matches 197; Conservative 31; Mismatches 25; Indels 0; Gaps 0;
QY 1 LLIVDVGNTVTVLGVYHDGKLEYHMRKTEDEFGMLRLSPHSGIMPEQIDGI 60
Db 1 MIFVLDVGNTVTVLGVYHDGDELKHHMRKTEDEFGIMTKALNLHVGLQSDIDGI 60
QY 61 IISVVPIMFALERMCTKRYFHIEPQIVGPMKGTGLNIRYDNPKREVAGDRIVNAVAIHL 120
Db 61 IISVVPIMFALERMCTKRYFHIEPQIVGPMKGTGLNIRYDNPKREVAGDRIVNAVAIHL 120
QY 121 YGNPLIVDFTATTATTCYIDENKQYMGALAPGITISFEALXSRAPKPRIETRPDII 180
Db 121 YGNPLIVDFTATTATTCYIDENKQYMGALAPGITISFEALXSRAPKPRIETRPDII 180
QY 121 YGNPLIVDFTATTATTCYIDENKQYMGALAPGITISFEALXSRAPKPRIETRPDII 180
Db 121 YGNPLIVDFTATTATTCYIDENKQYMGALAPGITISFEALXSRAPKPRIETRPDII 180

OY 181 GKTVSAMOSGILFGYVGVEGIVRKMKQAKODLKVATGGLAPLIANESDIDIVDPF 240
 DB 181 GKTVSAMAGILGIVGVEGIVSRMKAKSPVPKVIATGGLASLIASESNVIDIVDPF 240
 OY 241 LTLKGLLELYERN 253
 DB 241 LTLKGLLELYERN 253

RESULT 2

OY9KH5 PRELIMINARY: PRT: 254 AA.

AC OY9KH5: 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein BH0086.
 GN BH0086.
 OS Bacillus halodurans.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillales;
 NC Bacteria: Bacillus.
 NCBI_TaxID=86665;

RA [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001507; BAB03805.1; -
 DR InterPro: IPR004619; Baf.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMS: TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 254 AA; 27907 MW; 0E3B557BA7DAC176 CRC64;

Query Match 73.9%; Score 987; DB 16; Length 254;

Best Local Similarity 72.3%; Pred. No. 8,6e-80; Indels 0; Gaps 0;

Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

OY 1 LLLIVDVGNNTVLYGYHGGKLEYHWRIRTSRKTEDEFGMLIRSLFDHSGLMFEQIDGI 60
 DB 1 MLIVIDVGNNTVLYGYOBTLLVHWRIRTSRKTEDEFGMLIRSLFDHSGLMFEQIDGI 60
 OY 61 IISVVPPIIMFALERMCTYFHEIPQIVGPMKTGINKYDNPKREVGADRIYNAVAIHL 120
 DB 61 VISSVPPMMFSLQCKKRYFHTVPMIIGPIKTGLINKYDNPKREVGADRIYNAVAIHL 120
 OY 121 YGNPLIVDVGATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETTRPDNI 180
 DB 121 YGPAIVDVGATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETTRPDNI 180
 OY 181 GKTVSAMOSGILFGYVGVEGIVRKMKQAKODLKVATGGLAPLIANESDIDIVDPF 240
 DB 181 GNTTIDSMOSGIFGYVSGVDGVKMKAKQAESEPVVATGGLAKLIGTESRTIDIVDPF 240
 OY 241 LTLKGLLELYERN 253
 DB 241 LTLKGLLELYERN 253

RESULT 3

OY9AC5 PRELIMINARY: PRT: 259 AA.

AC OY9AC5: 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein lmo0221.
 GN lmo0221.

OS Listeria monocytogenes.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 NC Listeria.
 NCBI_TaxID=1639;

RA [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-E / SEROVAR 1/2A;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Clauser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Chardot A., Chetouani F., Couve E., de Darvar A., Deloux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
 RA Entlan K.-D., Fslhi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordisk G., Novella S., de Padlos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
 RT "Comparative genomics of Listeria species."
 RL Science 294:849-852(2001).
 DR EMBL: AL591974; CAD00748.1; -
 DR MEROPS: M41.009; -
 DR ListList: LMO00221; -
 DR InterPro: IPR004619; Baf.
 DR Pfam: PF03309; Bvg_acc_factor; 1.
 DR TIGRFAMS: TIGR00671; baf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 259 AA; 28187 MW; F106B049D80025B7 CRC64;

Query Match 63.8%; Score 852; DB 16; Length 259;

Best Local Similarity 63.4%; Pred. No. 9,1e-68; Indels 0; Gaps 0;

Matches 161; Conservative 37; Mismatches 56; Indels 0; Gaps 0;

OY 1 LLLIVDVGNNTVLYGYHGGKLEYHWRIRTSRKTEDEFGMLIRSLFDHSGLMFEQIDGI 60
 DB 1 MLIVIDVGNNTVLYGYOBTLLVHWRIRTSRKTEDEFGMLIRSLFDHSGLMFEQIDGI 60
 OY 61 IISVVPPIIMFALERMCTYFHEIPQIVGPMKTGINKYDNPKREVGADRIYNAVAIHL 120
 DB 61 VISSVPPMMFSLQCKKRYFHTVPMIIGPIKTGLINKYDNPKREVGADRIYNAVAIHL 120
 OY 121 YGNPLIVDVGATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETTRPDNI 180
 DB 121 YGPAIVDVGATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETTRPDNI 180
 OY 181 GKTVSAMOSGILFGYVGVEGIVRKMKQAKODLKVATGGLAPLIANESDIDIVDPF 240
 DB 181 GKSTVSMAOSGIFGYVSGVDGVKMKAKQAESEPVVATGGLAKLIGTESRTIDIVDPF 240
 OY 241 LTLKGLLELYERN 254
 DB 241 LTLKGLLELYERN 254

RESULT 4

OY9F54 PRELIMINARY: PRT: 259 AA.

AC OY9F54: 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein lln0253.
 GN lln0253.
 OS Listeria innocua.
 OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 NC Listeria.
 NCBI_TaxID=1642;
 RA [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / SEROVAR 6A;
 RX PubMed=11679669;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,

RESULT 5	PRELIMINARY:	PRT:	273 AA.
097EB4			
097EB4			
097EB4			
01-OCT-2001 (TREMBLrel. 18, Created)			
01-OCT-2001 (TREMBLrel. 18, Last sequence update)			
01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
Predicted transcriptional regulator, homolog of Bvg accessory factor.			
CAC3200.			
Clostridium acetobutylicum.			
Bacteria; Firmicutes; Bacilli; Clostridium group; Clostridia; Clostridiales; Clostridaceae; Clostridium.			
NCBI_TaxID=1488;			
[1]			
SEQUENCE FROM N.A.			
STRAIN=ATCC 824 / DSM 792 / VKM B-1787;			
MEDLINE=21359325; PubMed=11466286;			
Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hill J., Wolf Y.I., Ratusov R.L., Sabatthe F., Doucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;			
"Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum."			
J. Bacteriol. 183:4823-4838(2001).			
EMBL: AE007815; AAK81136.1; -.			
InterPro: IPR004619; Baf.			
InterPro: IPR000515; BPD, transp.			

[illegible]

```

OY 181 GKNTVSAMOSGILFGYGVGVEGIVKRMKQAKODKVIATGGLANESDPCIDVDPF 240
DB 181 NKRTVASMOSGILFGYGVGVEGIVKRMKQAFAPSAVYATGTFAMNAMESKIDTVNEM 240
OY 241 LTKGLIELIYERNR 254
DB 241 LTKGLIELIYERNR 254

RESULT 7
OY 08XHL5 PRELIMINARY: PRT: 259 AA.
AC 08XHL5;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Hypothetical protein CPE2468.
GN CPE2468.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
NCBI_Taxid=1502;
OX NCBI
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohnami K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL: AP003194; BAB82174.1;
DR InterPro: IPR004619; Baf.
DR InterPro: IPR000515; BPD.transp.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRfams: TIGR00671; Baf.1.
DR PROSITE: PS00402; BPD_TRANS-INN_MEMBER; UNKNOWN_1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 259 AA; 28819 MW; 0D5FEA3B7A145E10 CRC64;
SO

Query Match 55.2%; Score 737; DB 16; Length 259;
Best Local Similarity 56.2%; Pred. No. 1.5e-57;
Matches 145; Conservative 50; Mismatches 59; Indels 4; Gaps 1;

OY 1 LLLVIVGNNTVLYGYHDKLEYHMRIFETSRKTEDEFGMILRSLFDHSGLMFEQIDGI 60
DB 1 MLLIDVGNNTVLYGIDNKEYIASWRISTDSTKTSDEYSIOVQQLFNOAKLNPDEVEGI 60
OY 61 IISVVPPIWFALEKRCCTKFEHIEPOIVGPMKTGTLNIRKYNPREVGADRIYNAVAIHL 120
DB 61 IISVVPPIWFALEKRCCTKFEHIEPOIVGPMKTGTLNIRKYNPREVGADRIYNAVAIHL 120
OY 121 YGNPLIVVDFGATATTCYIDENKQYMGALPGITISTEALYSRAKLPRIETRPDNI 180
DB 121 YGNPLIVVDFGATATTCYIDENKQYMGALPGITISTEALYSRAKLPRIETRPDNI 180
OY 121 HKKMMIITDEGATATTCATFCEKDYLGNICPGIISADALFERAKLPRIETLEKPSVI 180
DB 121 HKKMMIITDEGATATTCATFCEKDYLGNICPGIISADALFERAKLPRIETLEKPSVI 180
OY 181 GKNTVSAMOSGILFGYGVGVEGIVKRMKQAKODKVIATGGLANESDPCID 236
DB 181 GKNTVSAMOSGILFGYGVGVEGIVKRMKQAKODKVIATGGLANESDPCID 236
OY 181 CKNTVTSQAGIITGYIGKVEYIVKRMKEMMDLGEKEPVLATGGLAKLYSETDVIDE 240
DB 181 CKNTVTSQAGIITGYIGKVEYIVKRMKEMMDLGEKEPVLATGGLAKLYSETDVIDE 240
OY 237 VDFELTKGLIELIYERNR 254
DB 237 VDFELTKGLIELIYERNR 254
OY 241 VDKLITLGLIELIYERNR 258
DB 241 VDKLITLGLIELIYERNR 258

RESULT 8
OY 09X8N6 PRELIMINARY: PRT: 265 AA.
AC 09X8N6;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
SO

```

```

DE Hypothetical protein SC03380.
GN SC03380 OR SCE94.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
NCBI_Taxid=1902;
OX NCBI
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Khasht H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
Mol. Microbiol. 21:77-96(1996).
RN (4)
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M45;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Radnolovich E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrett B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL: AL049628; CAB40880.1;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRfams: TIGR00671; Baf.1.
KW Hypothetical protein.
SO SEQUENCE 265 AA; 28203 MW; E34FA2ABBA946969 CRC64;

Query Match 49.1%; Score 655; DB 16; Length 265;
Best Local Similarity 51.0%; Pred. No. 3.1e-50;
Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;

OY 1 LLLVIVGNNTVLYGYHDKLEYHMRIFETSRKTEDEFGMILRSLFDHSGLMFEQ---- 56
DB 1 MLLIDVGNNTVLYGIDNKEYIASWRISTDSTKTSDEYSIOVQQLFNOAKLNPDEVEGI 60
OY 57 IDGIISSVVPPIWFALEKRCCTKFEHIEPOIVGPMKTGTLNIRKYNPREVGADRIYNAVAIHL 115
DB 57 IDGIISSVVPPIWFALEKRCCTKFEHIEPOIVGPMKTGTLNIRKYNPREVGADRIYNAVAIHL 115
OY 61 IDGIISSVVPPIWFALEKRCCTKFEHIEPOIVGPMKTGTLNIRKYNPREVGADRIYNAVAIHL 120
DB 61 IDGIISSVVPPIWFALEKRCCTKFEHIEPOIVGPMKTGTLNIRKYNPREVGADRIYNAVAIHL 120
OY 116 AALHLGNPLIVVDFGATATTCYIDENKQYMGALPGITISTEALYSRAKLPRIETRPDNI 175
DB 116 AALHLGNPLIVVDFGATATTCYIDENKQYMGALPGITISTEALYSRAKLPRIETRPDNI 175
OY 121 AAELVIGPAIYVDFGATATTCATFCEKDYLGNICPGIISADALFERAKLPRIETLEKPSVI 180
DB 121 AAELVIGPAIYVDFGATATTCATFCEKDYLGNICPGIISADALFERAKLPRIETLEKPSVI 180
OY 176 PDNIGKNTVSAMOSGILFGYGVGVEGIVKRMKQAKODKVIATGGLANESD 232
DB 176 PDNIGKNTVSAMOSGILFGYGVGVEGIVKRMKQAKODKVIATGGLANESD 232
OY 181 PRSVIGKNTVSAMOSGILFGYGVGVEGIVKRMKQAKODKVIATGGLANESD 240
DB 181 PRSVIGKNTVSAMOSGILFGYGVGVEGIVKRMKQAKODKVIATGGLANESD 240
OY 233 CIDVDFELTKGLIELIYERNR 253
DB 233 CIDVDFELTKGLIELIYERNR 253
OY 241 VIDHEPWLTKGLIELIYERNR 261
DB 241 VIDHEPWLTKGLIELIYERNR 261

RESULT 9

```


09A6Z1
ID 09A6Z1 PRELIMINARY; PRT; 261 AA.
AC 09A6Z1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria: Proteobacteria: alpha subdivision; Caulobacter group;
OC Caulobacter.
NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; Pubmed=11259647;
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
Pollock I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AEO05867; MAK23910.1; -;
DR TIGR: CC1935; -;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRfams: TIGR00671; Baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C1960D7B0714E95 CRC64;

Query Match 43.2%; Score 577; DB 16; Length 261;
Best Local Similarity 44.4%; Pred. No. 2.7e-43;
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;

QY 1 LLLVIVDGNNTVGLVYHDKLEHYHMRLETSRHKTEDEFGMLRSLPDSGLMPEQIDGI 60
DB 2 MLLAIEGNGNTMFALHIGASVWQVRSATSESTRADEYVWLSQGLSGRAIDAV 61
QY 61 IISVVPPIFALERMCTKFFHIEPQIVGPMKTKGLINKITDKNEVAGADRIYANVAIHL 120
DB 62 IISVVPQSIFFNLNLSRFRFNEPVIYGENAKLIDIVRIEKPEADRLVNAIGAMV 121
QY 121 YGNPLIVDGTATTCYIDENKQYMGATAPITISTEALYSAAKLPRIETRP--DN 178
DB 122 YPGPLVVIDSGTATTFIDIVAADGAFEGGILAPGINLSMQLHEAAKLPRIADQRPAGNR 181
QY 179 IICKNTVSAMQSGILFGVGVGEIVKRMKQAKQDLKVIATGGLAPLANESDCIDIVD 238
DB 182 IYGDVYSAMQSGVFWGVIYSLIEGLVARIKAREPMTVIATGVSALFEGATSIDHFD 241
QY 239 PELTLKLELIYERNR 255
DB 242 SDLTIRCLLEIYRRNTI 258

RESULT 10
Q9CD56
ID 09CD56 PRELIMINARY; PRT; 274 AA.
AC 09CD56;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein ML0232.
GN ML0232.
OS Mycobacterium leprae.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
OC Actinomycetales: Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=TN;
RX MEDLINE=21128732; Pubmed=11234002;
RA Cole S.T., Elgmeler K., Parkhill J., James K.D., Thomson N.R.,
Wheeler P.R., Honore N., Garner T., Churcher C., Harris D.,
Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
Holtroff S., Hornsby T., Jagsels K., Lacroix C., McLean J., Moulé S.,
Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus.";
RT Nature 409:1007-1011(2001).
RL Nature 409:1007-1011(2001).
DR EMBL: AL583917; CAC29740.1; -;
DR Leproma: ML0232; -;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRfams: TIGR00671; Baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 29421 MW; 1C2E735BDEC78765 CRC64;

Query Match 38.9%; Score 519.5; DB 16; Length 274;
Best Local Similarity 41.0%; Pred. No. 3.7e-38;
Matches 109; Conservative 53; Mismatches 89; Indels 15; Gaps 5;

QY 1 LLLVIVDGNNTVGLVY----HDKLEHYHMRLETSRHKTEDEFGMLRSLPDSGLMPEQ 56
DB 1 MLLAIDVNRHTTVGLSGSKHAKVVOQMRIRRESEVTADELALIIDGLGDS---ER 57
QY 57 IDGIISVVPPIFALERMCTKFFHIEPQIVGPMKTKGLINKITDKNEVAGADRIYANV 115
DB 58 LAGAAALSTVPSVLEHVRIMIDQYWPSPHVLIPPEVGTGPIPLVDPKREVGADRIYVCL 117
QY 116 AAILXGNPLIVDGTATTCYIDENKQYMGATAPITISTEALYSAAKLPRIETRP 175
DB 118 AAFHKEQAAALVDFGSSICVDVVSANGFEFGAIAQVOVSDAAARSALRLVELAR 177
QY 176 PDNITCKNTVSAMQSGILFGVGVGEIVKRMKQAKO---DL---KVITGGLAPLIA 228
DB 178 PRSVGKNTVBCMQAGVVFAGLVGDVIGMRDVEFSSDGLNRAVAVATGHPAPLL 237
QY 229 NESDCIDIVDFELTLKLELIYERNR 254
DB 238 PELHTVDHYDRHLLTGLRLVFEERNR 263

RESULT 11
O06282
ID 006282 PRELIMINARY; PRT; 272 AA.
AC 006282;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 29.3 kDa protein (transcriptional activator, putative, Baf family).
GN RV3600C OR MTCY07H7B.22 OR MT3706.
OS Mycobacterium tuberculosis.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
OC Actinomycetales: Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroff S.,
Hornsby T., Jagsels K., Krogh A., McLean J., Moulé S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the


```

QY 61 IISVVPPT---MFLERKCTKYEHIIEQIVGPGMKTGILNIKNDPKKEGADRIYNAYA 116
DB 64 VLSSVAPPVGENVALALKR---HFMDAFAVASAENLPVTELDIRGSGVADRLCNLFG 119
QY 117 AILHYG--NPLIVDFGTATTTCYIDENKQYMGALAPGTTSTELYSRAAKLPRIET 174
DB 120 AEKYLGLDVAAYVDGTSTNEFVYGRGRFLGLATGAQVSADLFAFAAKLPRIITLQ 179
QY 175 RPDNIIGKNTVASMOSGILFYGVEGIVKRMKQAKODLKVATGCLAPLIANESDCI 234
DB 180 APETAIGKNTVHALQSGLVGEAEMVDGLRIRIARLPEAANAVATGFSRTVGICQET 239
QY 235 DIVDFPLTKGL-ELIYENRV 255
DB 240 DYDETLRLGLVELWASRSEY 261

```

RESULT 14

```

Q9WZY5 PRELIMINARY: PRT: 246 AA.
Q9WZY5:
DB 01-NOV-1999 (TREMBLrel. 12, Created)
DB 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DB 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein TM0883.
CN TM0883.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Helt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher C.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:329-329(1999).
DR EMBL: AE001754; AAD35964.1;
DR TIGR: TM0883;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
DR TIGRfams: TIGR00671; Baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 246 AA; 27154 MW; 9E0309AD462CF266 CRC64;

```

Query Match

Best Local Similarity 30.4%; Score 405.5; DB 16; Length 246;
Matches 100; Conservative 42; Mismatches 88; Indels 25; Gaps 9;

```

QY 1 LLLVIVGNTNVLGVYHDGKLEYHMRKTSRHKTDEDFGMLRSLFDH---SGLMFPQ 56
DB 1 MYLLVIVGNTNHSYSLIEDGKTFRRMRLSTGCVQTEDE-----LFSHLHPLLGDAKRE 53
QY 57 IDGIITSVVPIIMFALERCTKYEHIIEQIVGPGMKTGL---NIKYDNPKEVGADRIYN 113
DB 54 IKIGICVAVPPTQNTVIERSKYFHSI--IIVKAKNGCVKWNV--NPSEVGADRVAN 109
QY 114 AVAAHLYGNPLIVDFGTATTTCYIDENKQYMGALAPGTTSTELYSRAAKLPRIET 173
DB 110 VVAFVKEYGKNGIITIDMGATVDLV-VNGSEGGAILPFGFMVHSLFRGTRAKLPLEV 168
QY 174 TRPDNIIGKNTVASMOSGILFYGVEGIVKRMKQAKODLKVATGCLAPLIANESDC 233
DB 169 KPADFVNGVQTEENIRLGVNGSVYALLEGIGRIK-EVYGDLPVLTGGOSKIY---KDM 224
QY 234 I--DIVDFPLTKGL 246
DB 225 IKHEIFDEDLTIKGV 239

```

RESULT 15

```

O32514 PRELIMINARY: PRT: 212 AA.
O32514:
DB 01-JAN-1998 (TREMBLrel. 05, Created)
DB 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DB 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to Bacillus subtilis.
OS Desulfovibrio vulgaris (strain Miyazaki).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MIYAZAKI;
RA Kitamura M., Konishi T., Kawashiri K., Ohashi K., Kishida Y.,
RA Kohno K., Akutsu H., Kumagai I., Nakaya T.;
RT "Sequence analyses of two ferredoxin genes and their flanking regions
RT from Desulfovibrio vulgaris (Miyazaki F).";
RL J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).
DR EMBL: AB005550; BAA21476.1;
DR InterPro: IPR004619; Baf.
DR Pfam: PF03309; Bvg_acc_factor; 1.
SQ SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;

```

Query Match

Best Local Similarity 25.4%; Score 338.5; DB 2; Length 212;
Matches 77; Conservative 40; Mismatches 83; Indels 7; Gaps 5;

```

QY 3 LLLVIVGNTNVLGVYHDGKLEYHMRKTSRHKTDEDFGMLRSLFDH---SGLMFPQ 62
DB 6 LLEDIGNTNKKIGIAVETAVLTSVLPDTPGQTTDSIGRLLEVRHAGLGRADVACVVA 65
QY 63 SSVVPIIMFALERCTKYEHIIEQIVGPG-MKTGILNKYDNPKEVGADRIYNAAHLY 121
DB 66 SSVVPGVNPILIRACERILY-RKLFPADIDALPLDNRYERAEVGADRLVAAYARRLY 124
QY 122 GNP--LIVDFGTATTTCYIDENKQYMGALAPGTTSTELYSRAAKLPRI--ETRPD 177
DB 125 PGRSLVSVDFGTATFDCV-EGCAVLGGLICPGVLSAGALSRAKLPRI-SLEVEDS 183
QY 178 NIKGNTVASMOSGILFYGVEGIV 204
DB 184 PVIGRSTTSLNKGFTFGFAMTEGVL 210

```

Search completed: April 21, 2003, 13:42:05
Job time : 87 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: April 21, 2003, 13:38:26 ; Search time 45 Seconds

(without alignments)
551.171 Million cell updates/sec

Title: US-09-813-453A-2

Perfect score: 1335
Sequence: 1 LLLVIVGNTNIVLGYVHIDG.....PFLTKGLIELIERNRNGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1114	83.4	233	2	S66100 conserved hypotet
2	987	73.9	254	2	P83660 hypothetical prote
3	852	63.8	259	2	AF1102 conserved hypotet
4	847	63.4	259	2	AF1164 conserved hypotet
5	775	58.1	273	2	E97293 probable transcrip
6	655	49.1	265	2	T36391 hypothetical prote
7	577	43.2	261	2	B87489 transcription acti
8	519.5	38.9	274	2	H86937 conserved hypotet
9	494.5	37.0	272	2	A70855 conserved hypotet
10	428	32.1	262	2	E75516 conserved hypotet
11	405.5	30.4	246	2	D72320 conserved hypotet
12	327.5	24.5	273	2	D71326 conserved hypotet
13	271.5	20.3	262	2	F70165 conserved hypotet
14	170	12.7	276	2	A12292 conserved hypotet
15	163	12.2	257	2	S75559 hypothetical prote
16	154.5	11.6	229	2	E70465 hypothetical prote
17	150	11.2	592	2	B81009 B1ra protein/Bvg a
18	134.5	10.1	242	2	A82637 conserved biotin-la
19	133	10.0	248	2	H83111 conserved hypotet
20	129.5	9.7	267	2	I40327 hypothetical prote
21	117	8.8	224	2	A99571 bat protein - Bord
22	116.5	8.7	223	2	G71887 conserved hypotet
23	109.5	8.2	223	2	F64627 hypothetical prote
24	109.5	8.2	209	2	H81382 hypothetical prote
25	104.5	7.8	597	2	B69251 probable electon
26	95	7.1	467	2	AE3142 hypothetical prote
27	95	7.1	520	2	G98145 probable aminotran
28	93	7.0	1014	2	H71602 protein with Dnad

30	91.5	6.9	828	2	G82583 bifunctional aspar
31	90.5	6.8	4452	1	YGBS62 gramacidin S synth
32	88	6.6	256	2	H83002 imidazoleglycerol-
33	87.5	6.6	287	2	D90540 glucokinase (gluco
34	87	6.5	336	2	AE0625 dihydroorotate deh
35	86.5	6.5	205	2	AC0612 anaerobic dimethyl
36	86.5	6.5	205	2	AG0680 probable dimethyl
37	86.5	6.5	243	2	D64300 (R)-2-hydroxygluta
38	86	6.4	336	1	S13824 dihydroorotate oxi
39	85	6.4	336	1	DEECD0 dihydroorotate oxi
40	85	6.4	336	2	E90757 dihydro-orotate de
41	85	6.4	336	2	C85621 dihydro-orotate de
42	84.5	6.3	317	2	E95792 dihydro-orotate de
43	84	6.3	443	1	JC5298 protein kinase (EC
44	83.5	6.3	205	2	B85615 hypothetical prote
45	83.5	6.3	205	2	D90751 hypothetical prote

ALIGNMENTS

RESULT 1
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C:Accession: S66100; E69740
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A:Title: Systematic sequencing of the 180 kDa region of the Bacillus subtilis ch
A:Reference number: S65967; MID:96051385; PMID:7584024
A:Accession: S66100
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-233 <OGA>
A:Molecule type: DNA
A:Cross-references: EMBL:D26185; NID:9467326; PIDN:BA05305.1; PID:d1005847; PID:9467
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerston, P.T.; Entian, K.D.; Erlington, J.; Fábreg, C.; Ferrati,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, F
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinc
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sc
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchida
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MID:98044033; PMID:9384377
A:Accession: E69740
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-233 <KUN>
A:Cross-references: GB:299104; GB:AL009126; NID:92632267; PIDN:CA11846.1; PID:el182C
A:Experimental source: strain 168
C:Genetics:
A:Gene: yacB
A:Start codon: TTG
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c
Query Match 83.4%; Score 1114; DB 2; Length 233;
Best Local Similarity 99.5%; Pred. No. 3.6e-88;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

[illegible]

RESULT 5

Probable transcription regulator, homolog of Bvg accessory factor [Imported] - Clostridium
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_rev:10114 #text_change 30-Sep-2001
 C:Accession: E97293
 R:Rolling, J., Breton, G., Omelchenko, M.V., Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J.; Bailey, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J:Entry: D183_4833-4838, 2001
 J:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium
 J:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: E97293
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-273 <KUR>
 A:Cross-references: CB:AE001437; PTD:NAK81136.1; PTD:G15026270; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3200
 A:Superfamily: Streptomyces coelicolor hypothetical protein SCE94_31c

Query Match	Score	DB 2	Length	273
58.18	775			

best local similarity 30.4%; pred. NO. 3.4e-59;
Matches 146; Conservative 55; Mismatches 54; Indels 4; Gaps 1;

Oy 1 LLLVLDVGNTNIVGLVYXHDGKLEXYHRIETSRHKTEDEFGMLTSLRFDHSGLMFEQIDGI 600
 ::|||:||||| |||::|| |::| :||::: |||
Dd 12 VILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADYEGLIQVMNLFGDDKLDPFLVEGV 711

```

Oy 61 ISSVPPINFALEERCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIYNAVAAIL 120
      ||||| :|:| | | | | :|||:|||||:|||||:|||||:
Db 72 ISSVPPNIMVSLHEMIRKRYFKINPLVVGPGIKTCINIKYDNPKEVGADRIYNAVAHET 131

```

```

09 121 YGNPLIVDEGTATYCYIDENKQYMGCAIAPGITISTEALYŠRAKLPRIETRPDNI 180
    |  ||:|||||:| : || :||| || :|||:| :|||:| :|
Db 132 YKRSLLIIDEGTATTECAVRENGDYLGCAICPGIKVSSEALFEKAKLPYRELKPAVAI 191

```

```

Oy 181 GKNTVSAMQSCILEFGVQVEIEVKRKMKQAKODK----VIATGGLAPLIANESDCIDI 238
    |||:::||||:|||| ||::: :::: |::| ||: |: :|
Ob 192 CKNTISSISGSIYGIQVRYIVERMEELQEGEKEPLVATGTGLAKLTISEAKNDV 251

```

```
Oy      237  VDPFLTLKGLIYERNRV  255
          ::|||||:| :|||:|
Db      252  INFPLTEGLRIYEKNRV  270
```

RESULT 6
T36391

hypothetical protein SCE94_31c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36391
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A:Reference number: Z21573
A:Accession: T36391
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-265
A:Cross-references: EMBL:AL049628; PTDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94_31c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE94_31c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94_31c

Query Match	49.1%;	Score 655;	DB 2;	Length 265;
Best Local Similarity	51.0%;	Pred. No. 1e-48;		
Matches 133;	Conservative 46;	Mismatches 74;	Indels 8;	Gaps 3

```
1 LILVIDGNNITVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ---- 566
```

57 IDGIITSSVVPIMFALERMCKTKYFHEIPQI-VGPGMKTGLNKKYDNPKEVGADRIYNNAV 115

61 IDGAIACAVPSVLHELHEVTRRRYGGDPAVLVEPVKTGVPILTDHPKEVGADRINAV 120
116 AAHHYGNPLIVDFGTATTCYIDENKQYMGGAIAIPCITISTEALYSRAAKLPRIETR 175
11: ||||| ||||| : :||| ||||| : ||| : ||| :

176 PDNIGKNTVASMQGILGYVGCGIVKRMKMQA---KQDKLVATGSLAPLANESD 232

181 PRSV1GKNIVEAMQSGIYVGFAGVDGVNRMARELADDDVTVIATGSLAPMVLGESS 240
233 CIDIVDPFLTKGLELIYERN 253
II :I:III II I:III

241 VIDEHEPWLTLMGLRLVYERN 261
RESULT 7

Species

3: Accession: B87489

Nielman, W.C.; Feidolyun, I.V.; Paulsen, L.T.; Nelson, K.E.; Eissen, J.; Heidegger, B.; Laub, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ermolaeva, M.I.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, J.C.; Natl Acad Sci U.S.A. 98, 4136-4141, 2001

Accession: B87489
 Status: preliminary
 Reference number: A87249; MUID:21173698; PMID:1129647
 Complete genome sequence of candidate circulant

```

molecule_type: DNA
residues: 1-261 <STO>
Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
Genetics:

```

Query Match .43.2% Score 577; DB 2; Length 261;
;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

1 LLLVIVDGNITNVILGVYHDDKLEHYHWRIETSRHKTEDEFGMLRLSLFDHSGLMFEQIDCI 60
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1

2 MLAI^{EQ}GNTMMF^{AI}HDGASV^{WA}Q^{NR}SAT^{ES}RT^{AD}EYV^{WL}SQ^{LS}MG^{LG}RA^{ID}AV 61
61 IISV^{VP}PI^{MA}L^{ER}M^{CT}KY^{FI}EP^{IQ}IV^{GP}GM^{TG}LN^{IK}DN^{PK}EV^{GA}DR^{IV}NA^{VA}I^{HL} 12

62 IISVVPOSIFNLRLNLSRRYFNEVPLVIGENAKLIGDIVRIEKRPSEAGADRLVNAIGAAV 121
121 YGNPLIVDEGTATTCYIDENKQYMGAIAPGITISTEALYSRAAKLPRIEITRP--DN 178

122 YGPIPLVIDSGTATTEDIVAADGAEEGGLIAPGINLSMQLHEAAKLPRIATQIPAGNR 18
179 IICKNTVSAMSGILFCYGVQVEGIIVKRMKQAKQDLKVIATGGLAPLIANESDCIDIYD 23

182 IVCIDTVSAMQSGVFWGYISLIEGLVARIKAEGERMPTVIATGTVASLFEGATDSIDHFD 24
239 PELTKGLELIYERNRV 255

242 SDLTIRGLLEIYRRTI 258
b6
b7C
b7D
RESULT 8

RESULT 8

H86937
 conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: H86937
 R:Coile, S.T.; Elgiel, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
 R.; Davies, R.M.; Devlin, K.; Duthey, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sg
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A86909; PMID:21128732; PMID:11234002
 A:Accession: H86937
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-274 <STO>
 A:Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: ML0232
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.9%; Score 519.5; DB 2; Length 274;
 Best Local Similarity 41.0%; Pred. No. 4.5e-37;
 Matches 109; Conservative 53; Mismatches 89; Indels 15; Gaps 5;
 1 LLLIVDGNNTVIGVY---HDGKLEYHRIETSRKTEDEFGMILRSIFDHSGLMFEQ 56
 1 MLAIIDVNRTHVYVGLSSKKEHAKVQOMRITSEVYADELALIIIDLGSDS---ER 57
 57 IDGIISVVPPIPMALERNCKRYFHEIPQI-VGPMKGTGLNIRKYNPKREVADRIYNAV 115
 58 LAGAAALSTVPSVLEHVRITMLDQYMPSPVHLEPGVIRGIPILVDNPKREVADRIYNAV 117
 116 AATHLYGNPLIVDREGTATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIETR 175
 118 AARHKGQAALIVDFGSSICVDVYSAKGEFLGALAPGVQVSSDAARSAALRYELAR 177
 176 PDNIITKNTVSAMSGILFGYGOVEGIYKRMKQAKQ--DL---KVATGGLAPLI 228
 178 PRSVGKNTECKMGAGVGFAGLVGRMQRQDVEFGSDIGNVAVATHTAPLL 237
 229 NESDCIDIVDPFLTKGLIELYERNR 254
 238 PELHTVHDYDRLHTLGLRLVERNR 263

RESULT 9

A:0935
 hypothetical protein RV3600c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70955
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 R.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 R.; Raftery, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Stares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; PMID:98295987; PMID:9634230
 A:Accession: A70955
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-272 <COL>
 A:Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113976
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3600c
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 37.0%; Score 494.5; DB 2; Length 272;
 Best Local Similarity 38.8%; Pred. No. 6.3e-35;
 Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

1 LLLIVDGNNTVIGVY---HDGKLEYHRIETSRKTEDEFGMILRSIFDHSGLMFEQ 56
 1 MLAIIDVNRTHVYVGLSSKKEHAKVQOMRITSEVYADELALIIIDLGSDS---ER 57
 57 IDGIISVVPPIPMALERNCKRYFHEIPQI-VGPMKGTGLNIRKYNPKREVADRIYNAV 115
 58 LAGAAALSTVPSVLEHVRITMLDQYMPSPVHLEPGVIRGIPILVDNPKREVADRIYNAV 117
 116 AATHLYGNPLIVDREGTATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIETR 175
 118 AARHKGQAALIVDFGSSICVDVYSAKGEFLGALAPGVQVSSDAARSAALRYELAR 177
 176 PDNIITKNTVSAMSGILFGYGOVEGIYKRMKQAKQ--DL---KVATGGLAPLI 228
 178 PRSVGKNTECKMGAGVGFAGLVGRMQRQDVEFGSDIGNVAVATHTAPLL 237
 231 SDCIDIVDPFLTKGLIELYERNR 253
 238 PELHTVHDYDRLHTLGLRLVERNR 260

RESULT 10

E75516
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
 C:Accession: E75516
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
 M.; Shen, M.; Vamathevan, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; PMID:20036896; PMID:10567266
 A:Accession: E75516
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-262 <MHI>
 A:Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AA110040.1; PID:g645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0461
 A:Map position: 1
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 32.1%; Score 428; DB 2; Length 262;
 Best Local Similarity 37.4%; Pred. No. 3e-29;
 Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;

1 LLLIVDGNNTVIGVYH-DGKLEYHRIETSRKTEDEFGMILRSIFDHSGLMFEQIDG 60
 6 LLAVIDGNNTVIGLADAGALHTWRIRIRNEMLPDIALQLHGLTLAGADIPR--AA 63
 61 IISVVPPI---MFLERNCKRYFHEIPQI-VGPMKGTGLNIRKYNPKREVADRIYNAV 116
 64 VLSVAPVPGENYALDKR---HFMIDAFVSAEMLPDYVELDIPGSGVADRLCNLGG 119
 117 AATHLYGNPLIVDREGTATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIETR 174
 120 AEXYLLGDLVAVVDDGTSTNPFVYGRGRFRFLGILATGQVADALFRAAKLPRIITLQ 179
 175 RPDNIITKNTVSAMSGILFGYGOVEGIYKRMKQAKQKDKVATGGLAPLIANSDDCI 234
 180 APTAIGAKNTVHALQSGLVFGYAEWVDGLIRIRIAPLPGBAFAVAVATGFSRTVQIGOEI 239
 235 DIYDPEFLTKGL-ELIYERNR 255
 240 DYEDELTLTKGLVLEWASREV 261

RESULT 11

D72320
 conserved hypothetical protein - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: D73320
 A: Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.
 Nature 399, 323-329, 1999
 A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
 A:Reference number: A72200; MUID:99287316; PMID:10360571
 C:Accession: D73320
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-246 <ARN>
 A:Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AD35964.1; PID:g498141
 A:Experimental source: strain MS8
 C:Genetics:
 A:Gene: TM0883
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.4%; Score 405.5; DB 2; Length 246;
 Best Local Similarity 39.2%; Pred. No. 2.4e-27;
 Matches 100; Conservative 42; Mismatches 88; Indels 25; Gaps 9;
 Db 1 LLLVDGNTNTVGLVYHDGKLEIWMRIETSRHKTDEDFGMLNSLFDH---SGLMFEQ 56
 1 LLLVDGNTNTVGLVYHDGKLEIWMRIETSRHKTDEDFGMLNSLFDH---SGLMFEQ 56
 Db 1 MLLVDGNTNTVGLVYHDGKLEIWMRIETSRHKTDEDFGMLNSLFDH---SGLMFEQ 53
 57 IDGIISVVPIMFALERMCTKFFHIEPQIVGPMKGL---NKKYDNPEKVGADRYN 113
 54 IKGIIVASVVPIMFALERMCTKFFHIEPQIVGPMKGL---NKKYDNPEKVGADRYN 109
 Db 114 AVAAIHILGNPLIVDFGATATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIE 173
 110 VVAIFKEKNGKIIIDMTATTVLV-VNCSYEGGAILPGFFMVHSLFRTAKLPLEV 168
 Db 174 TRPDNIIGNKTVSAMQSGILFEGYGOVGYIVKRMKQAKODLYATAGLAPLANSDC 233
 169 KPADVVCKOREENIRIGVNGSVYALEGITGRIRK-EVYGDLPVLTGGOSKIV---KDM 224
 Db 234 I--DIVPFLTKGL 246
 225 IKHEIFEDLTIKGV 239

RESULT 12
 D71326
 Conserved hypothetical protein TP0431 - syphilis spirochete
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
 C:Accession: D71326
 A: Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, D.; Peterson, J.; Kierulff, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDevitt, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A:Reference number: A71250; MUID:98332770; PMID:9665876
 C:Accession: D71326
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-273 <COL>
 A:Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332272
 A:Experimental source: strain Nichols
 C:Genetics:
 A:Gene: TP0431
 C:Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 24.5%; Score 327.5; DB 2; Length 273;
 Best Local Similarity 29.9%; Pred. No. 1.3e-20;
 Matches 75; Conservative 64; Mismatches 105; Indels 7; Gaps 4;
 Db 2 LVIDGNTNTVGLVY--HDGK--LEIWMRIETSRHKTDEDFGMLNSLFDHSGLMFEQ 57
 1 MLLIDGNSHVYFGIOGNGKVCVAREFLRLAPDARKTODEYSLILHALCERAGVGRASL 60

Query 58 DGIISVVPIMFALERMCTKFFHIEPQIVGPMKGLTNKYDN--KEVGADRYNAV 115
 61 RDAFISVVPIMFALERMCTKFFHIEPQIVGPMKGLTNKYDN--KEVGADRYNAV 120
 Db 116 AAHILGNPLIVDFGATATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIE 175
 121 AAHILGNPLIVDFGATATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIE 160
 Db 176 PDNIIGNKTVSAMQSGILFEGYGOVGYIVKRMKQAKODLYATAGLAPLANSDC 235
 181 PDNIIGNKTVSAMQSGILFEGYGOVGYIVKRMKQAKODLYATAGLAPLANSDC 239
 Db 236 IVDVPLTKGL 246
 240 PIDAQLTSLG 250

RESULT 13
 F70165
 Conserved hypothetical protein BB0527 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: F70165
 A: Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Laibinre, R.; Whiston, D.; Peterson, J.; Kierulff, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDevitt, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 A>Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 C:Accession: F70165
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-262 <KLE>
 A:Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g2688431
 A:Experimental source: strain B31

Query Match 20.3%; Score 271.5; DB 2; Length 262;
 Best Local Similarity 27.3%; Pred. No. 8.2e-16;
 Matches 70; Conservative 57; Mismatches 108; Indels 21; Gaps 5;
 Db 3 LVIDGNTNTVGLVYHDGKLEIWMRIETSRHKTDEDFGMLNSLFDHSGLMFEQ 62
 9 LVIDGNTNTVGLVYHDGKLEIWMRIETSRHKTDEDFGMLNSLFDHSGLMFEQ 62
 Db 63 SSVPIMFALERMCTKFFHIEPQIVGPMKGLTNKYDN--NPKK---VGADRYNA 114
 63 SSVPIMFALERMCTKFFHIEPQIVGPMKGLTNKYDN--NPKK---VGADRYNA 116
 Db 115 VVAIHILGNPLIVDFGATATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIE 173
 117 VVAIHILGNPLIVDFGATATTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIE 176
 Db 174 TRPDNIIGNKTVSAMQSGILFEGYGOVGYIVKRMKQAKODLYATAGLAPLANSDC 233
 177 TRPDNIIGNKTVSAMQSGILFEGYGOVGYIVKRMKQAKODLYATAGLAPLANSDC 236
 Db 234 IVDVPLTKGL 249
 237 EPIFNIHLTVESGRL 252

RESULT 14
 A12292
 Hypothetical protein alr3896 (imported) - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: A12292
 A: Kaneo, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Irigun, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, M.; DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2292
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-276 <KOR>
A:Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17133030; GSPDB:GM00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3896

Query Match 12.7%; Score 170; DB 2; Length 276;
Best Local Similarity 24.2%; Pred. No.4.5e-07;
Matches 66; Conservative 50; Mismatches 115; Indels 42; Gaps 10;

OY 3 LVIDVGNNTWLYGVYHDKLEHYMRIE-----TSRHKTDEDFGMLRSLFDHSGLM 53
|::||: |:: ||: |:: ||: |:: ||: |
15 LALEIGNSRHLMAFLMGSSLEFMDTELPESVIQQLNGSTKLVEOSEKEIR----FT 70

OY 54 FPOID-----GIISVVPPIMPALERMCRTKYFHIEPQIVGPCMKTGLNIKNYPREV 107
|::||: |:: ||: |:: ||: |:: ||: |
71 PPPLPACOPLPFLFIASVPDQTVLME---NYLNVRVTITLD--QIPLNNIPT--LG 120

OY 108 ADRIYNANAHIILGNPLIIVDFGPATTCYCIDEKKOMGAIAPGITISTEALYSRAAK 167
|:: ||: |:: ||: |:: ||: |:: ||: |
121 IDRALMALMGMSGFPAVLVIDAGALTFFTAADGGKNLVGAILPVGLOPSLSGOQTQC 180

OY 168 LPRIETRPDN---IGKNTVSAMOGSILFGYGVEGIVARMKWQA-QDLKIYAIFGL 223
|::||: |:: ||: |:: ||: |:: ||: |
181 LQOVEEAKAISLPFRFAINTTEAIQSIVYTLLIAGRPFTE--EWLSLFPDGKVAIKGSD 238

OY 224 APLIANESDCID-----YDPFLTTLKGLELI 249
|:: ||: |:: ||: |:: ||: |
DB 239 RIIILNVYLQALYPDLAARLIVERPNLIWFWMQPI 271

RESULT 15
S75559
hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:accession: S75559
R:kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O., K.; Okumura, S.; Shimpo, S.; Takuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, S.
DNA Res. 3, 109-136, 1996
A:title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.

A:Reference number: S74322; MUID:97061201; PMID:8905231
Accession: S75559
Status: nucleic acid sequence not shown; translation not shown
Molecule type: DNA
A:Residues: 1-257 <KAN>
A:Cross-references: EMBL:D90911; GB:AB001339; MID:g1653083; PIDN:BA18120.1; PID:d101885
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:start codon: GTG

Query Match 12.2%; Score 163; DB 2; Length 257;
Best Local Similarity 28.0%; Pred. No.1.6e-06;
Matches 58; Conservative 42; Mismatches 79; Indels 28; Gaps 10;

OY 60 TIISVVPIMPALERMCTRYPHI-EPOIVGPMKTKGLINIKYDN-PREVGADRIVANAAVA 117
|::||| |:: |:: |:: |:: |:: |
DB 64 IMLASVPE-----QTEVMRYQPRL-----TLKNPLVNLYPSGISDALAGLT 110

OY 118 IHLYGNPLIIVDFGPATTCYCIDEKKOMGAIAPGITISTEALYSRAAKLPRIETR-- 175
|:: ||: |:: ||: |:: ||: |:: ||: |
DB 111 GLTYGFPOLVDGTALTITGFDDKRKVAGALLPGIGLOLATLGRNALPRLENDOLF 170

OY 176 --PDNIICKNTVASMOGSILFGYGYGVEGIVARMKWQA-QDLKIYATGSG---LAPLIA 228
|:: ||: |:: ||: |:: ||: |:: ||: |
DB 171 ELIPDR-WALDTPSAIFSGVAYGVALAQSYIQ--DWCKLEPGAAMVTTGDGKITILGFELK 227

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 13:28:21 : Search time 38 Seconds
(Without alignments)
904.702 Million cell updates/sec

Title: US-09-813-453A-2

Perfect score: 1335
Sequence: 1 LLLVLDVGNNTNVLGVYHDG.....PFLTKLELIERNRYGSV 258

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1335	100.0	258	AAU01243	B. subtilis novel
2	1335	100.0	258	AAU01149	Bacillus subtilis
3	1114	83.4	233	AAU01163	Pantothenate kinase
4	1046	78.4	258	AAU01172	Pantothenate kinase
5	1034	77.5	262	AAU01170	Pantothenate kinase
6	987	73.9	254	AAU01171	Pantothenate kinase
7	852	63.8	259	AB847661	Listeria monocytogenes
8	794.5	59.5	256	AAU01175	Pantothenate kinase
9	756	56.6	255	AAU01154	Geobacter sulfurreducens
10	655	49.1	265	AAU01151	Streptomyces coelicolor

11	646.5	48.4	250	AAU01150	Clostridium acetobutylicum
12	577	43.2	260	AAU01173	Pantothenate kinase
13	537	40.2	258	AAU01153	Rhodospirillum rubrum
14	515	38.6	219	AAU01176	Pantothenate kinase
15	494.5	37.0	272	AAU01152	Mycobacterium tuberculosis
16	493.5	37.0	272	AAU01155	Mycobacterium tuberculosis
17	428	32.1	262	AAU01155	Mycobacterium tuberculosis
18	405.5	30.4	246	AAU01156	Mycobacterium tuberculosis
19	338.5	25.4	212	AAU01177	Thermotoga maritima
20	327.5	24.5	273	AAU01157	Pantothenate kinase
21	319	23.9	257	AAU01174	Treponema pallidum
22	271.5	20.3	262	AAU01158	Pantothenate kinase
23	207	15.5	244	AAU01168	Borrelia burgdorferi
24	203	15.2	241	AAU01179	Pantothenate kinase
25	163	12.2	249	AAU01182	Pantothenate kinase
26	163	12.2	257	AAU01180	Synechocystis sp. PCC 6803
27	154.5	11.6	229	AAU01159	Aquifex aeolicus
28	154	11.5	249	AAU01178	Pantothenate kinase
29	151	11.3	455	AAU01159	Neisseria gonorrhoeae
30	151	11.3	455	AAU01157	Neisseria gonorrhoeae
31	151	11.3	460	AAU01167	Pantothenate kinase
32	151	11.3	592	AAU01167	Neisseria meningitidis
33	151	11.3	592	AAU01167	Neisseria meningitidis
34	150	11.2	389	AAU01169	Neisseria meningitidis
35	150	11.2	455	AAU01169	Neisseria meningitidis
36	150	11.2	592	AAU01169	Neisseria meningitidis
37	150	11.2	592	AAU01169	Neisseria meningitidis
38	150	11.2	592	AAU01169	Neisseria meningitidis
39	150	11.2	592	AAU01169	Neisseria meningitidis
40	150	11.2	592	AAU01169	Neisseria meningitidis
41	150	11.2	592	AAU01169	Neisseria meningitidis
42	134.5	10.1	242	AAU01180	Pantothenate kinase
43	133	10.0	248	AAU01164	Pantothenate kinase
44	129.5	9.7	267	AAU01164	Pantothenate kinase
45	114.5	8.6	189	AAU01164	Neisseria meningitidis

ALIGNMENTS

RESULT 1
ID AAU01243 standard; Protein: 258 AA.
XX
AC AAU01243;
XX
DT 18-JUL-2001 (first entry)
XX
DE B. subtilis novel pantothenate kinase encoded by the gene coxX.
XX
KW Pantothenate kinase; coxX; pantothenate biosynthesis; vitamin B5;
KW nutritional supplement; panto-compound; pantoate.
XX
OS Bacillus subtilis.
XX
PN WO200121772-A2.
XX
PD 29-MAR-2001.
XX
PE 21-SEP-2000; 2000WO-US25993.
XX
PR 21-SEP-1999; 99US-0400494.
PR 07-JUN-2000; 2000US-0210072.
PR 28-JUL-2000; 2000US-0221836.
PR 24-AUG-2000; 2000US-0227860.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS.
XX
PI Yocum RR, Patterson TA, Hermann T, Pero JG;
DR WPI: 2001-218644/22.
DR N-PSDB; AAS00984;
XX

PT New recombinant microorganism which overexpress a *Bacillus subtilis*
 PR pantothenate biosynthetic enzyme, useful for the high yield production
 PR of panto-compounds such as pantothenate and pantoate
 XX
 PS Example 14; Fig 23; 292pp; English.

CC The sequence represents a novel B. subtilis pantothenate kinase (encoded
 CC by gene cox), an enzyme of the pantothenate biosynthetic pathway.
 CC Pantothenate, also known as vitamin B5, is used as a nutritional
 CC supplement in mammals and humans. The invention concerns methods of
 CC producing recombinant microorganisms overexpressing at least one *Bacillus*
 CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods
 CC of producing them are useful for producing a panto-compound such as
 CC pantothenate or pantoate, which is a nutritional requirement for
 CC livestock and humans. The methods are also useful for the identification
 CC of pantothenate kinase modulators. Panto-compounds are produced at a
 CC significantly higher yield than prior art methods and can be produced
 CC independent of the need to feed precursors which decreases expense.

XX
 SQ Sequence 258 AA:

Query Match 100.0%; Score 1335; DB 22; Length 258;
 Best Local Similarity 100.0%; Pred. No. 8.2e-142;
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLIVDVGNTNTVLYGYHDKLEYHWRITSRKTEDEFGMLRSFPHSGLMFQIDGI 60
 DB 1 LLLIVDVGNTNTVLYGYHDKLEYHWRITSRKTEDEFGMLRSFPHSGLMFQIDGI 60
 QY 61 IISVVPPIMFALERMCTYFHEIPQIVGPKMTGLINIKYDNPKEVGADRIYNAVAIHL 120
 DB 61 IISVVPPIMFALERMCTYFHEIPQIVGPKMTGLINIKYDNPKEVGADRIYNAVAIHL 120
 QY 121 YGNPLIVDFGATATYCYIDENKQYMGAIAPGITTSTALYSRAAKLPRIETTRPDNII 180
 DB 121 YGNPLIVDFGATATYCYIDENKQYMGAIAPGITTSTALYSRAAKLPRIETTRPDNII 180
 QY 181 GKNTVSAMOSGILFGYGVGVEGIVRKMKQAKODLKVIATGGLAPLIANESDCIDIVDF 240
 DB 181 GKNTVSAMOSGILFGYGVGVEGIVRKMKQAKODLKVIATGGLAPLIANESDCIDIVDF 240
 QY 241 LFLKGLIELIYERNRGSV 258
 DB 241 LFLKGLIELIYERNRGSV 258

RESULT 2
 AAU91149
 AAU91149 standard; Protein; 258 AA.

AAU91149;

XX 05-JUN-2002 (first entry)

XX *Bacillus subtilis* pantothenate kinase Coax.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX *Bacillus subtilis*.

OS WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX Yocum RR, Patterson TA;
 PI

XX WPI: 2002-269358/31.
 DR N-PSDB: ABR54168.

PT Identifying potential antibiotic or antimicrobial agent, comprises
 PR contacting composition comprising pantothenate kinase (Coax) protein
 PR with test compound and identifying inhibitor of the Coax protein

PS Claim 10; Page 67-68; 128pp; English.

XX The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.

XX
 SQ Sequence 258 AA:

Query Match 100.0%; Score 1335; DB 23; Length 258;
 Best Local Similarity 100.0%; Pred. No. 8.2e-142;
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLIVDVGNTNTVLYGYHDKLEYHWRITSRKTEDEFGMLRSFPHSGLMFQIDGI 60
 DB 1 LLLIVDVGNTNTVLYGYHDKLEYHWRITSRKTEDEFGMLRSFPHSGLMFQIDGI 60
 QY 61 IISVVPPIMFALERMCTYFHEIPQIVGPKMTGLINIKYDNPKEVGADRIYNAVAIHL 120
 DB 61 IISVVPPIMFALERMCTYFHEIPQIVGPKMTGLINIKYDNPKEVGADRIYNAVAIHL 120
 QY 121 YGNPLIVDFGATATYCYIDENKQYMGAIAPGITTSTALYSRAAKLPRIETTRPDNII 180
 DB 121 YGNPLIVDFGATATYCYIDENKQYMGAIAPGITTSTALYSRAAKLPRIETTRPDNII 180
 QY 181 GKNTVSAMOSGILFGYGVGVEGIVRKMKQAKODLKVIATGGLAPLIANESDCIDIVDF 240
 DB 181 GKNTVSAMOSGILFGYGVGVEGIVRKMKQAKODLKVIATGGLAPLIANESDCIDIVDF 240
 QY 241 LFLKGLIELIYERNRGSV 258
 DB 241 LFLKGLIELIYERNRGSV 258

RESULT 3
 AAU91163
 AAU91163 standard; Protein; 233 AA.

AAU91163;

XX 05-JUN-2002 (first entry)

XX Pantothenate kinase (Coax) #1.

XX Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.

XX *Bacillus subtilis*.

OS WO200216601-A2.

XX 28-FEB-2002.

XX 24-AUG-2001; 2001WO-US26531.

XX 24-AUG-2000; 2000US-227860P.

XX 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX
 PA

XX Vocum RR, Paterson TA;
 PI
 XX
 DR WPI: 2002-269358/31.
 DR N-PSDB: ABK54169.
 XX
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein
 XX
 PS Disclosure; Page 81-82; 128pp; English.
 CC
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 CC
 SQ Sequence 233 AA;
 Query Match 83.4%; Score 1114; DB 23; Length 233;
 Best Local Similarity 99.5%; Pred. No. 5,9e-117;
 Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLLVIDGNTNTVGVYHDKLEYHWRLETSRHKTEDEFGMILRSLEFDSGLMFEQIDGI 60
 DB 1 MLTVIDGNTNTVGVYHDKLEYHWRLETSRHKTEDEFGMILRSLEFDSGLMFEQIDGI 60
 QY 61 IISVVPPIPMFALERMCKTFYFIEPOIVGPGMKTGLNIRKNDPKEVGADRIYNAVAIHL 120
 DB 61 IISVVPPIPMFALERMCKTFYFIEPOIVGPGMKTGLNIRKNDPKEVGADRIYNAVAIHL 120
 QY 121 YGNPLIVDFGATTTTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIETRPDNI 180
 DB 121 YGNPLIVDFGATTTTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIETRPDNI 180
 QY 181 GKNYVSAMOSGILFEGYGOVEGIVRKMKQAOKD 214
 DB 181 GKNYVSAMOSGILFEGYGOVEGIVRKMKQAOKD 214
 QY 181 GKNYVSAMOSGILFEGYGOVEGIVRKMKQAOKD 214
 DB 181 GKNYVSAMOSGILFEGYGOVEGIVRKMKQAOKD 214
 RESULT 4
 ID AAU91172
 ID AAU91172 standard; Protein; 258 AA.
 AC AAU91172;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Pantothenate kinase (Coax) #10.
 XX
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX
 OS Bacillus stearothermophilus.
 OS
 PN WO200216601-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26531.
 XX
 PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX
 PI Vocum RR, Paterson TA;

XX
 DR WPI: 2002-269358/31.
 DR N-PSDB: ABK54193.
 XX
 XX Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein
 XX
 PS Claim 10; Page 101-102; 128pp; English.
 CC
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 CC
 SQ Sequence 258 AA;
 Query Match 78.4%; Score 1046; DB 23; Length 258;
 Best Local Similarity 78.3%; Pred. No. 3.2e-109;
 Matches 198; Conservative 32; Mismatches 23; Indels 0; Gaps 0;
 QY 1 LLLVIDGNTNTVGVYHDKLEYHWRLETSRHKTEDEFGMILRSLEFDSGLMFEQIDGI 60
 DB 1 MFLVDGNTNTVGVYHDKLEYHWRLETSRHKTEDEFGMILRSLEFDSGLMFEQIDGI 60
 QY 61 IISVVPPIPMFALERMCKTFYFIEPOIVGPGMKTGLNIRKNDPKEVGADRIYNAVAIHL 120
 DB 61 IISVVPPIPMFALERMCKTFYFIEPOIVGPGMKTGLNIRKNDPKEVGADRIYNAVAIHL 120
 QY 121 YGNPLIVDFGATTTTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIETRPDNI 180
 DB 121 YGNPLIVDFGATTTTCYIDENKQYMGALAPGTTISTEALYSRAKLPRIETRPDNI 180
 QY 181 GKNYVSAMOSGILFEGYGOVEGIVRKMKQAOKD 214
 DB 181 GKNYVSAMOSGILFEGYGOVEGIVRKMKQAOKD 214
 QY 241 LTKGLELYERN 253
 DB 241 LTKGLELYERN 253
 RESULT 5
 ID AAU91170
 ID AAU91170 standard; Protein; 262 AA.
 AC AAU91170;
 XX
 DT 05-JUN-2002 (first entry)
 XX
 DE Pantothenate kinase (Coax) #8.
 XX
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX
 OS Bacillus anthracis.
 OS
 PN WO200216601-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26531.
 XX
 PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.
 PA (OMNI-) OMNIGENE BIOPRODUCTS INC.

XX Yocum RR, Patterson TA;
 XX WPI: 2002-269358/31.
 DR N-PSDB: ABR54191.
 XX
 PT Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 PS Claim 8: Page 98-99; 128pp; English.
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 SQ Sequence 262 AA;
 Query Match 77.5%; Score 1034; DB 23; Length 262;
 Best Local Similarity 75.1%; Pred. No. 7.3e-108; Indels 0; Gaps 0;
 Matches 190; Conservative 34; Mismatches 29;
 QY 1 LLVLDVGNNTVYGVYDGLKLEYHMRIRTSRHKTEDEFGMTLSLFDHSGLMFEQIDGI 60
 DB 1 MIFVLIDVGNNTVYGVYDGLKLEYHMRIRTSRHKTEDEFGMTLSLFDHSGLMFEQIDGI 60
 QY 61 IISVVPPIEMFALERMCTKTFHIEPQIVGPMKGTGLNRYDNPKRYGADRIVNAAIAHL 120
 DB 61 IVSSVVPPIEMFALERMCTKTFHIEPQIVGPMKGTGLNRYDNPKRYGADRIVNAAIAHL 120
 QY 121 YGNPLIVDFGATATYCYIDENKQYWGAIAPGITISTEALYSRAAKLPRIETIRPDNII 180
 DB 121 YGSPPLIVDFGATATYCYIDENKQYWGAIAPGITISTEALYSRAAKLPRIETIRPDNII 180
 QY 181 GKNVTSAMQSGILFGYGOVEGYIKRMKQAKODLKVIATGGLAPLIANESDCIDIVDPF 240
 DB 181 GKNVTSAMQSGILFGYGOVEGYIKRMKQAKODLKVIATGGLAPLIANESDCIDIVDPF 240
 QY 241 LTLKGLIYERN 253
 DB 241 LTLKGLIYERN 253
 SUIT 6
 AA091171
 ID AA091171 standard; Protein: 254 AA.
 AC AA091171;
 XX
 DT 05-JUN-2002 (first entry)
 DE Pantothenate kinase (Coax) #9.
 XX
 KW Pantothenate kinase; Coax; antibiotic; antimicrobial;
 KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 OS Bacillus halodurans.
 XX
 PN WO200216601-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US26531.
 XX
 PR 24-AUG-2000; 2000US-227860P.
 PR 20-MAR-2001; 2001US-0813453.

XX (OMNI-) OMNIGENE BIOPRODUCTS INC.
 PA Yocum RR, Patterson TA;
 XX WPI: 2002-269358/31.
 DR N-PSDB: ABR54192.
 XX
 PT Identifying potential antibiotic or antimicrobial agent, comprises
 PT contacting composition comprising pantothenate kinase (Coax) protein
 PT with test compound and identifying inhibitor of the Coax protein -
 PS Claim 10: Page 100; 128pp; English.
 CC The invention describes assays for identifying a (potential) antibiotic
 CC comprising contacting an assay composition comprising a pantothenate
 CC kinase (Coax) protein with a test compound, and determining the ability
 CC of the test compound to inhibit the activity of the Coax protein, an
 CC essential enzyme for the production of coenzyme A. Coax protein is a
 CC valuable target for identifying bactericidal compounds. Coax modulating
 CC agents can be used in an infectious animal model to determine the
 CC efficacy, toxicity, or side effects of treatment with such an agent. This
 CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 CC described in the invention.
 SQ Sequence 254 AA;
 Query Match 73.9%; Score 987; DB 23; Length 254;
 Best Local Similarity 72.3%; Pred. No. 1.4e-102; Indels 0; Gaps 0;
 Matches 183; Conservative 33; Mismatches 37;
 QY 1 LLVLDVGNNTVYGVYDGLKLEYHMRIRTSRHKTEDEFGMTLSLFDHSGLMFEQIDGI 60
 DB 1 MIFVLIDVGNNTVYGVYDGLKLEYHMRIRTSRHKTEDEFGMTLSLFDHSGLMFEQIDGI 60
 QY 61 IISVVPPIEMFALERMCTKTFHIEPQIVGPMKGTGLNRYDNPKRYGADRIVNAAIAHL 120
 DB 61 IVSSVVPPIEMFALERMCTKTFHIEPQIVGPMKGTGLNRYDNPKRYGADRIVNAAIAHL 120
 QY 121 YGNPLIVDFGATATYCYIDENKQYWGAIAPGITISTEALYSRAAKLPRIETIRPDNII 180
 DB 121 YGSPPLIVDFGATATYCYIDENKQYWGAIAPGITISTEALYSRAAKLPRIETIRPDNII 180
 QY 181 GKNVTSAMQSGILFGYGOVEGYIKRMKQAKODLKVIATGGLAPLIANESDCIDIVDPF 240
 DB 181 GKNVTSAMQSGILFGYGOVEGYIKRMKQAKODLKVIATGGLAPLIANESDCIDIVDPF 240
 QY 241 LTLKGLIYERN 253
 DB 241 LTLKGLIYERN 253
 RESULT 7
 ABB47661
 ID ABB47661 standard; Protein: 259 AA.
 AC ABB47661;
 XX
 DT 05-FEB-2002 (first entry)
 DE Listeria monocytogenes protein #365.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.

(INSP) INST PASTEUR.

WPI; 2002-010914/01.

Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment and prevention of *Listeria* and related bacterial infections, and related polypeptides -

Claim 6; SEQ ID No 366; 192pp; French

The present invention relates to the genome sequence of *Listeria monocytogenes* EGD-e (see AB403041). The genome sequence and fragments of it are useful for selecting probes and primers for detecting genes in *L. monocytogenes* and related organisms, and for studying genetic polymorphisms and other genomes. The present sequence is a protein expressed from the genome sequence of the present invention. Proteins expressed from the genome sequence are useful for raising specific antibodies. Identification of *L. monocytogenes* and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of Vitamin B12. The genome sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate *L. monocytogenes*-related diseases. In addition, the genome sequence and proteins encoded by it are useful in pharmaceutical and vaccines compositions for the treatment or prevention of infections by *L. monocytogenes* and related organisms.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

SQ Sequence 259 AA;

Query Match	63.8%	Score 852;	DB 23;	Length 259;
Best Local Similarity	63.4%	Pred. No. 2.4e-87;		
Matches 161:	Conservative 37;	Mismatches 56;	Indels 0;	Gaps 0

machines 101; conservative 37; Mismatches 56; Indels 0; Gaps 0

```

1 LLLVDBGNTNVLGVYHDGKLEEHWRIRIETSRKHTEDEFGMLRLSLFDHSGLMFEQIDGI 60
:::|||||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
1 MTHUDICGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG

```

61 IISVVPIMFALERMCTKYEHEPOLVPGMKGTGLNIKYDNPKEVGADRITNAAVAATH. 120

```

61 IISVPPIMHMETMCVRYFNIRPLIVGPGIKIGLNKVDNPREIGSDRIANVAASEE 120

```

```

121 YGNPbVWDBSTATTCYIDENKQMGCAIAPGTTISTELYSRAAKLPREITRPDNI 180
||:|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
VCEDEITITDIDGDMEECTGTRF-----:|||||:|||||:|||||:|||||

```

181 GKNTVSAMQSGILFGYVGQVEGIVKRMKQWQAKODKVIATGGAPLIANESDCTIDYDPE 240

181 GKSTVSSMQAGIFGYGCGCEGIIAEMKQSNNSPVVATGGLARRITTEKSSAVDIDPF 240

```

241 LTLGELLEIYERNR 254
      | | | | | | | | | |
241 LMTKCEIYVDDYV 254

```

AAU91175 standard: Protein: 256 aa

AAU91175;

De Pantothenate kinase (Coax) #13
XX

KW Pantothenate kinase; Coax; antibiotic; antimicrobial
KW Pantothenate kinase modulator; Coax; antibiotic; antimicrobial

pallolinolate kinase modulator; coenzyme A; bactericidal compound

OS Clostridium difficile.
xy

PN WO200216601-A2

PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26531.

PR 24-AUG-2000; 2000US-227860P.

PR 20-MAR-2001; 2001US-0813453.
XX

PA (OMNI-) OMNIGENE BIOPRODUCTS INC
XX

PI Yocum RR, Patterson TA;
XY

DR WPI; 2002-269358/31.
DR N-PSDB; APK54105

Idos + 4 Enr + 2 Co + 247

PT contacting composition comprising pantothenate kinase (Coax) protein with test compound and a test effect.

With test compound and identifying inhibitor of the Coax protein -

PS Claim 6; Page 105; 128pp; English.
XX

The invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

SQ Sequence 256 AA;

Query Match	59.58;	Score 794.5;	DB 23;	Length 256
Best Local Alignment	60.00	803.00	23	256

Best Local Similarity 60.88; Pred. No. 7.2e-81;
Matches 155; Conservative 40; Mismatches 59; Indels 1; Cons 1

1 LLLVLDVGNINVLGVYHDGKLEHYMRLETSRHKTEDEFCMLRLSLDLHDHSGLMFEQIDGI 60

1 MLVVDVGNINMVLGIYKGDKLNVYWRILKTDREKTSDEYCGILISNEDYDNNISDIDV 60

[illegible]

121 YGNPIIVDEGTATVYCIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180

121 YGAPSLVDEGTATTCFAISEGGEYLLGTIAPGIRISSEALFQASKSLPREVELAKPGMTI 180

181 CKSTVYSAMSGIIITGVGVIKDKITSTIMKKEI.NCDDIVKVTATGCGIARTTASGMRSTVNDV 340

240 FTIKGLELIYERNR 254
I I I I I I I I I I

241 FIVEGRITTEKNQ 255

RESULT 9
W091154

AA091154 standard; Protein; 255 AA.

[illegible]

XX	AAU91151;
AC	05-JUN-2002 (first entry)
XX	Streptomyces coelicolor pantothenate kinase Coax.
DE	Pantothenate kinase; Coax; antibiotic; antimicrobial;
XX	pantothenate kinase modulator; coenzyme A; bactericidal compound.
KW	Streptomycetes coelicolor.
OS	MO200216601-A2.
XX	
PN	28-FEB-2002.
XX	
PD	24-AUG-2001; 2001WO-US26531.
XX	
PF	24-AUG-2000; 2000US-227860P.
XX	
PR	20-MAR-2001; 2001US-0813453.
XX	
PA	(OMNI-) OMNIGENE BIOPRODUCTS INC.
XX	
PI	Yocum RR, Patterson TA;
XX	
DR	WPI: 2002-269358/31.
XX	
PT	Identifying potential antibiotic or antimicrobial agent, comprising
PT	contracting composition comprising pantothenate kinase (Coax) protein
PT	with test compound and identifying inhibitor of the Coax protein -
PS	Claim 10; Page 69-70; 128pp; English.
XX	
CC	The invention describes assays for identifying a (potential) antibiotic
CC	comprising contracting an assay composition comprising a pantothenate
CC	kinase (Coax) protein with a test compound, and determining the ability
CC	of the test compound to inhibit the activity of the Coax protein, an
CC	essential enzyme for the production of coenzyme A. Coax protein is a
CC	valuable target for identifying bactericidal compounds. Coax modulating
CC	agents can be used in an infectious animal model to determine the
CC	efficacy, toxicity, or side effects of treatment with such an agent. This
CC	is the amino acid sequence of a pantothenate kinase (Coax) protein
CC	described in the invention.
SQ	Sequence 265 AA:
	Query Match 49.1%; Score 655; DB 23; Length 265;
	Best Local Similarity 51.0%; Pred. No. 4e-65;
	Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3
OY	1 LLLIVDVGNTNTVLGVYHDKLEYHWRIETSRHKTEDEFGMILRSLPFDSGLMFEO----- 56
	: : :: :: : : : : : : :
DB	1 MLTIDVGNHTVLGLDFGEDIEVHEWRISTDSRRRTADELAVLLQGDMGHPILLGDELGDG 60
	:: :: :: :: : : : : : : : : : : : : : :
OY	57 IDIIITSSVVPRLMFLERMCCTKVFHLEPOI-VGPGRKTGLNTIKYDNPKREVGADRIYNAY 115
	:: :: :: :: : : : : : : : : : : : : : :
DB	61 IDSIALCAVTPSVLHEIREVTNRITRGDAVYLVEPVKTSVPLTDHPKEVGADRILINAY 120
	:: :: :: :: : : : : : : : : : : : : : :
OY	116 AAHLHGKNPLIVDFGTATTTCYIDENKOYGALAPGITISTEALYSRAAKLRPIETR 175
	:: :: :: :: : : : : : : : : : : : : : :
DB	121 ANVELVGGPALIVDFGATTFDAVSARGEYIGVIAFGIEISVEALGVKAQLRKIEVAR 180
	:: :: :: :: : : : : : : : : : : : : : :
OY	176 PDNIIGKNIVSAMOSGILGFYGVQVEGVIYKRMMQA---KODKYLATGS LAPLINESD 232
	:: :: :: :: : : : : : : : : : : : : : :
DB	181 PRSYIGKNTVEAAGSGIVGYFAQVGVGNRMARELADDPDVYATATGSLAPVILESS 240
	:: :: :: :: : : : : : : : : : : : : : :
OY	233 CIDIVDPFLTKGDELIERN 253
	:: :: :: :: : : : : : : : : : : : : : :
DB	241 VIDHEHPWLTLMLRLVYERN 261
	:: :: :: :: : : : : : : : : : : : : : :

[illegible]

DT	XX	05-JUN-2002 (first entry)
DE	XX	Pantothenate kinase (Coax) #11.
XX	XX	Pantothenate kinase (Coax); antibiotic; antimicrobial;
KW	XX	pantothenate kinase modulator; coenzyme A; bactericidal compound.
OS	XX	Caulobacter crescentus.
XX	XX	W0200216601-A2.
PN	XX	28-FEB-2002.
PD	XX	24-AUG-2001: 2001MO-US26531.
PF	XX	24-AUG-2000: 2000US-227860P.
PR	XX	20-MAR-2001: 2001US-0813453.
XX	XX	(OMNI-) OMNIGENE BIOPRODUCTS INC.
PA	XX	Yocum RR, Patterson TA;
XX	XX	WPI, 2002-269358/31.
DR	XX	N-PSDB; ABR54194.
XX	XX	Identifying potential antibiotic or antimicrobial agent, comprises
PT	XX	contacting composition comprising pantothenate kinase (Coax) protein
PT	XX	with test compound and identifying inhibitor of the Coax protein
PS	XX	Claim 10; Page 102-103; 128pp; English.
CC	XX	The invention describes assays for identifying a (potential) antibiotic
CC	XX	comprising contacting an assay composition comprising a pantothenate
CC	XX	kinase (Coax) protein with a test compound, and determining the ability
CC	XX	of the test compound to inhibit the activity of the Coax protein, an
CC	XX	essential enzyme for the production of coenzyme A. Coax protein is a
CC	XX	valuable target for identifying bactericidal compounds. Coax modulating
CC	XX	agents can be used in an infectious animal model to determine the
CC	XX	efficacy, toxicity, or side effects of treatment with such an agent. This
CC	XX	is the amino acid sequence of a pantothenate kinase (Coax) protein
CC	XX	described in the invention.
SQ	XX	Sequence 260 AA:
Query Match	43.2%;	Score 577; DB 23; Length 260;
Best Local Similarity	44.4%;	Pred. No. 2.4e-56;
Matches 114; Conservative	50;	Mismatches 91; Indels 2; Gaps 1
QY	1	LELVIVDVGNTNVVLGYHDKGLKLEHYHRRITSRKHTDERGMLRLSFLDHSGLMFEDIDGI 60
DB	1	MLLAIEGQNTNMFALIHGASVVAQWRSATSTETRTADEYVVMLSQLSMQGLGFRADAV 60
QY	61	IIISVVPPIIMFALERMCKRYFHEIEPOIVGPKMKTGLINIVDNPKEVGADRIYNAVAIHL 120
DB	61	IIISVVSPOSIFLRLNLSRXYFNEVPELVYENAKKLGDIVAIERKPESEGADRLVNAIGAMV 120
QY	121	YGNPLIVVDGFGATATYCYIDENKQYMGCAIAIGITITSTEALYSRAAKLPRIEITRP--DN 178
DB	121	YGPPLVVIDSGATTFEDIVAAADGAEFGGIAPGINLMSQALHEAAKLPRIAIQRPACNR 180
QY	179	IIGKMTVASMSGIIEGVGQEVGIVIKRKKWQAKODLKVYIATGSLAPLIANSDDIDVD 238
DB	181	IYGTDTVASMSGVFWGIISLIEGLIVARIKAEERGEPMYVATGVAISLFECAFTSDIHFD 240
QY	239	PFLTIKGLLELIYERRRV 255
DB	241	SDLTIRGLLEIYRRNTI 257

XX AC AAU91153;
 XX 05-JUN-2002 (first entry)
 XX DT Rhodobacter capsulatus pantothenate kinase Coax.
 XX DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
 XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX OS Rhodobacter capsulatus.
 XX PN WO200216601-A2.
 XX PD 28-FEB-2002.
 XX PF 24-AUG-2001; 2001WO-US26531.
 XX PR 24-AUG-2000; 2000US-227860P.
 XX PR 20-MAR-2001; 2001US-0813453.
 XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX PI Yocum RR, Patterson TA;
 XX DR WPI; 2002-269358/31.
 XX DR
 XX XX
 XX PT Identifying potential antibiotic or antimicrobial agent, comprises
 XX PT contacting composition comprising pantothenate kinase (Coax) protein
 XX PT with test compound and identifying inhibitor of the Coax protein
 XX PS Claim 10; Page 71-72; 128bp; English.
 XX PS
 XX CC The invention describes assays for identifying a (potential) antibiotic
 XX CC comprising contacting an assay composition comprising a pantothenate
 XX CC kinase (Coax) protein with a test compound, and determining the ability
 XX CC of the test compound to inhibit the activity of the Coax protein, an
 XX CC essential enzyme for the production of coenzyme A. Coax protein is a
 XX CC valuable target for identifying bactericidal compounds. Coax modulating
 XX CC agents can be used in an infectious animal model to determine the
 XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
 XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 XX CC described in the invention.
 XX CC
 XX SQ Sequence 258 AA;
 XX
 XX Query Match 40.28; Score 537; DB 23; Length 258;
 XX Best Local Similarity 43.9%; Pred. No. 7.8e-52;
 XX Matches 112; Conservative 46; Mismatches 95; Indels 2; Gaps 2;

ID AAU91176 standard; Protein; 219 AA.
 XX AC AAU91176;
 XX 05-JUN-2002 (first entry)
 XX DT Pantothenate kinase (Coax) #14.
 XX DE Pantothenate kinase; Coax; antibiotic; antimicrobial;
 XX KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
 XX OS Dehalococcoides ethenogenes.
 XX PN WO200216601-A2.
 XX PD 28-FEB-2002.
 XX PF 24-AUG-2001; 2001WO-US26531.
 XX PR 24-AUG-2000; 2000US-227860P.
 XX PR 20-MAR-2001; 2001US-0813453.
 XX PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
 XX PI Yocum RR, Patterson TA;
 XX DR WPI; 2002-269358/31.
 XX DR N-PSDB; ABK54197.
 XX DR
 XX XX
 XX PT Identifying potential antibiotic or antimicrobial agent, comprises
 XX PT contacting composition comprising pantothenate kinase (Coax) protein
 XX PT with test compound and identifying inhibitor of the Coax protein
 XX PS Claim 10; Page 106-107; 128bp; English.
 XX PS
 XX CC The invention describes assays for identifying a (potential) antibiotic
 XX CC comprising contacting an assay composition comprising a pantothenate
 XX CC kinase (Coax) protein with a test compound, and determining the ability
 XX CC of the test compound to inhibit the activity of the Coax protein, an
 XX CC essential enzyme for the production of coenzyme A. Coax protein is a
 XX CC valuable target for identifying bactericidal compounds. Coax modulating
 XX CC agents can be used in an infectious animal model to determine the
 XX CC efficacy, toxicity, or side effects of treatment with such an agent. This
 XX CC is the amino acid sequence of a pantothenate kinase (Coax) protein
 XX CC described in the invention.
 XX CC
 XX SQ Sequence 219 AA;
 XX
 XX Query Match 38.6%; Score 515; DB 23; Length 219;
 XX Best Local Similarity 45.8%; Pred. No. 1.9e-49;
 XX Matches 97; Conservative 45; Mismatches 70; Indels 0; Gaps 0;

AC AAU91152;

DT 05-JUN-2002 (first entry)
 YY

Mycobacterium tuberculosis pantothenate kinase Coax

KW pantothenate kinase; CoAX; antidiabetic; antimicrobial;

pancreatic kinase modulator; coenzyme A; bactericidal compound

05 *Mycobacterium tuberculosis*.
xy

PN WO200216601-A2.

PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US26531.
VY

PR 24-AUG-2000; 2000US-227860P.
PB 20-MAR-2001 2001US-0012473

20-MAR-2001; 2001US-0813453.

OMNIGENE BIOPRODUCTS INC.

PL YOCUM KR, PATERSON TA;
XX

WPI; 2002-269358/31.

identifying potential antibiotic or antimicrobial agent, comprises contacting composition with at least one

will test compound and identifying inhibitor of the Coax protein -

claim 10; page 70-71; 128bpp; English.

comprising contacting an assay compo

one invention describes assays for identifying a (potential) antibiotic comprising contacting an assay composition comprising a pantothenate kinase (Coax) protein with a test compound, and determining the ability of the test compound to inhibit the activity of the Coax protein, an essential enzyme for the production of coenzyme A. Coax protein is a valuable target for identifying bactericidal compounds. Coax modulating agents can be used in an infectious animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. This is the amino acid sequence of a pantothenate kinase (Coax) protein described in the invention.

Sequence 272 AA;

Query Match	37.08;	Score 494.5;	DB 23;	Length 272
Best Local Similarity	38.88;	Score 503.47;	DB 23;	Length 272

matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4.

```

OY 1 LLLVLDGNTNNVLGYV----HDGLFETHYHRIETSRKTFDEGMYLRSLPDSHLMRFO 56
Db 1 MLLAIIVRNHHVVGCLLSGMRKHAAVVOQWRIETRESEVTDALALIDDLIEDS---ER 57
OY 57 IDGIITSSVVPRIIMFLFEMCQKYPHLEIPOI-VQPGMKJGLANKYONPREVGADRIYAV 116
Db 58 LRTGALSTVPSVSLHEHVRIMLDQYMPSPVPHILIEPKRISILPLVDNPEVGADRIIVCL 117
OY 116 AALHLGNPLCYIVDGTATFYCYIDENKQYMGCAIAPGITISTEALYSAAKLPIRIETR 175
Db 118 AAVDRRKAAIIVDGGSSICLVADYVSAKGEEFGGIAIAGVOVSDAAAASALRLVEIAR 177
OY 176 PDIIITKNYVSAMQSGILFEGYVGOESIVRBMK----WQAKODLKVIATGGLAPLIANE 230
Db 178 PSHVGNKNTVECMQAGAAVFGAGLVLDLIVGRIREDVSGFSVDHVAIVATGHTAPLLPE 233
OY 231 SDCIDIVPELTLTKGLEIYERN 253
Db 238 LHTVVDYDQHTLGLRLVYERN 260

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 13:39:21 ; Search time 30 seconds

(without alignments)
253.037 Million cell updates/sec

Title: US-09-813-453A-2

Perfect score: 1335
Sequence: 1 LLLVIDGNNTVLTGVYHDG.....PFLTLKGLIELYERRRVGSV 258

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.5	6.3	207	3 US-08-549-515-9	Sequence 9, Appl 1
2	79	5.9	342	4 US-08-096-181A-12	Sequence 12, Appl 1
3	79	5.9	342	4 US-08-096-181A-14	Sequence 14, Appl 1
4	79	5.9	342	5 PCT-US94-08326-12	Sequence 12, Appl 1
5	79	5.9	342	5 PCT-US94-08326-14	Sequence 14, Appl 1
6	79	5.9	361	4 US-08-096-181A-8	Sequence 8, Appl 1
7	79	5.9	361	5 PCT-US94-08326-8	Sequence 8, Appl 1
8	79	5.9	363	4 US-08-096-181A-10	Sequence 10, Appl 1
9	79	5.9	363	5 PCT-US94-08326-10	Sequence 10, Appl 1
10	79	5.8	923	3 US-08-936-135-6	Sequence 6, Appl 1
11	78	5.8	1012	1 US-07-944-943-2	Sequence 2, Appl 1
12	78	5.8	1012	1 US-08-219-262B-2	Sequence 2, Appl 1
13	78	5.8	1012	2 US-09-031-655-2	Sequence 2, Appl 1
14	78	5.8	1213	3 US-08-937-102-2	Sequence 2, Appl 1
15	78	5.8	3898	2 US-08-876-891-2	Sequence 2, Appl 1
16	78	5.8	3898	2 US-09-059-853-2	Sequence 2, Appl 1
17	78	5.8	3898	4 US-08-750-717-2	Sequence 2, Appl 1
18	77	5.8	627	4 US-09-071-035-452	Sequence 452, App
19	77	5.8	1313	4 US-09-071-035-450	Sequence 450, App
20	77	5.8	1313	4 US-09-116-473-4	Sequence 454, App
21	76	5.7	922	4 US-09-116-473-4	Sequence 4, Appl 1
22	75.5	5.7	443	4 US-09-594-193-9	Sequence 9, Appl 1
23	75.5	5.7	1215	6 US-09-134-001C-5319	Sequence 9, Appl 1
24	75	5.6	553	6 5310678-1	Patent No. 5310678
25	75	5.6	1012	3 US-08-219-262B-9	Sequence 9, Appl 1
26	75	5.6	1012	3 US-09-031-655-9	Sequence 9, Appl 1
27	74.5	5.6	457	4 US-09-134-001C-3838	Sequence 3838, Ap

28	74	5.5	413	4 US-09-147-938-2	Sequence 2, Appl 1
29	73	5.5	724	4 US-08-914-999-4	Sequence 4, Appl 1
30	73	5.5	2319	1 US-08-212-133A-8	Sequence 8, Appl 1
31	73	5.5	2319	1 US-08-474-503-6	Sequence 6, Appl 1
32	73	5.5	2319	2 US-08-670-707A-6	Sequence 6, Appl 1
33	73	5.5	2319	4 US-09-037-601-6	Sequence 6, Appl 1
34	73	5.5	2319	4 US-09-315-179-6	Sequence 6, Appl 1
35	73	5.5	2319	4 US-09-523-656-28	Sequence 28, Appl 1
36	73	5.5	2319	5 PCT-US94-13400-6	Sequence 6, Appl 1
37	72.5	5.4	396	2 US-08-878-989-16	Sequence 16, Appl 1
38	72.5	5.4	396	4 US-08-272-796-16	Sequence 16, Appl 1
39	72.5	5.4	396	4 US-09-344-700-2	Sequence 2, Appl 1
40	72.5	5.4	820	1 US-08-291-896-2	Sequence 2, Appl 1
41	72.5	5.4	820	2 US-08-485-278-2	Sequence 2, Appl 1
42	72.5	5.4	1148	4 US-09-106-075A-90	Sequence 90, Appl 1
43	72.5	5.4	1612	1 US-08-169-927-2	Sequence 2, Appl 1
44	72	5.4	314	4 US-09-134-001C-2983	Sequence 2983, Ap
45	72	5.4	361	3 US-09-032-372-12	Sequence 12, Appl 1

ALIGNMENTS

```
RESULT 1
US-08-549-515-9
: Sequence 9, Application US/08549515
: Patent No. 6054123
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M
: APPLICANT: Klein, Michel H
: TITLE OF INVENTION: Haemophilus Influenzae
: TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Slim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/549,515
: FILING DATE: 27-OCT-1995
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-522
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 207 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-549-515-9

Query Match      6.3%; Score 83.5; DB 3; Length 207;
Best Local Similarity 24.4%; Pred. No. 0.11;
Matches 47; Conservative 27; Mismatches 66; Indels 53; Gaps 11;

QY 15 GYHDKLEIHWRIETSRHKTEDFGMLRLSLDHSGLMPEQIDGITIISVPPPIFALE 74
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 52 GYHDKLEIHWRIETSRHKTEDFGMLRLSLDHSGLMPEQIDGITIISVPPPIFALE 74
||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 75 RMC--TKYPIHIEQIVGPKMTGLNIRYDNP-----EVGADRIYNAVAAILHYGNPLI 126
```

Db 97 DVCIGCRHCHMACPYCAP-----QYNETGHHMTKCGCYDRAEAGKPKVCESPLR 148
QY 127 VDFGATATYCYIDENKQYMG--GAIAPGITSTALYSRAKLPRIETRPDNIIGNKT 184
Db 149 ALDFGP-----IDELARKHGLAANAP-----RA--LPRAHHTK-NIVIKRN 188
QY 185 VSAMQSGLIFGY 197
Db 189 ANSRPTGDTGYL 201

RESULT 2

US-08-096-181A-12
Sequence 12, Application US/08096181A
Patent No. 6153406

GENERAL INFORMATION:

APPLICANT: Tai, Joseph Y.
APPLICANT: Soper, Jeffrey K.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level Expression,
Purification And Refolding Of The Outer Membrane Protein
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,181A
FILING DATE: 23-Jul-1993
CLASSIFICATION: 424

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-096-181A-12

Query Match 5.9%; Score 79; DB 4; Length 342;
Best Local Similarity 20.1%; Pred. No. 0.83; Mismatches 71; Indels 62; Gaps 10;
Matches 43; Conservative 38;

QY 28 IETSRHKT-----EDEFGLILRSLE-----DHSGLMFEQIDGIISSVPPIMFAL 73
Db 100 VKLGRATITADGITSADKEVGLNNSDYIPSGNTVGYTFKIDGLVIGA---NYLLAQ 156
QY 74 ERMCTKYFHIEPQIVGPKMTGLINIKYDNPKEVGADRIYNAVAAILHLYGNPLIVVDGTA 133
Db 157 KREGAKGENKRPNDKAGEVRIQ---EINNGIOVGAKYDANDIYAKIAYGR-----203
QY 134 TTYCY--IDENKQYMGAIAP-----GITISTEALYSRAKLP-----R 170
Db 204 TNYKYNESDHKQOLNGVLAITLGRPSDLGLVSLDSGYAKTKYKIKHEKRYVSPGFQ 263
QY 171 IEITRPDNIIG-----KNTV-----SAMQSGLIFG 195
Db 264 YEIMEDTIVYGNFKYERTSVDOGEKTRQAVLFG 297

RESULT 3

US-08-096-181A-14
Sequence 14, Application US/08096181A
Patent No. 6153406

GENERAL INFORMATION:

APPLICANT: Tai, Joseph Y.
APPLICANT: Pullen, Jeffrey K.
APPLICANT: Soper, Thomas S.
APPLICANT: Liang, Shu-Mei
TITLE OF INVENTION: A Method For The High Level Expression,
Purification And Refolding Of The Outer Membrane Protein
TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,181A
FILING DATE: 23-Jul-1993
CLASSIFICATION: 424

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-096-181A-14

Query Match 5.9%; Score 79; DB 4; Length 342;
Best Local Similarity 20.1%; Pred. No. 0.83; Mismatches 71; Indels 62; Gaps 10;
Matches 43; Conservative 38;

QY 28 IETSRHKT-----EDEFGLILRSLE-----DHSGLMFEQIDGIISSVPPIMFAL 73
Db 100 VKLGRATITADGITSADKEVGLNNSDYIPSGNTVGYTFKIDGLVIGA---NYLLAQ 156
QY 74 ERMCTKYFHIEPQIVGPKMTGLINIKYDNPKEVGADRIYNAVAAILHLYGNPLIVVDGTA 133
Db 157 KREGAKGENKRPNDKAGEVRIQ---EINNGIOVGAKYDANDIYAKIAYGR-----203
QY 134 TTYCY--IDENKQYMGAIAP-----GITISTEALYSRAKLP-----R 170
Db 204 TNYKYNESDHKQOLNGVLAITLGRPSDLGLVSLDSGYAKTKYKIKHEKRYVSPGFQ 263
QY 171 IEITRPDNIIG-----KNTV-----SAMQSGLIFG 195
Db 264 YEIMEDTIVYGNFKYERTSVDOGEKTRQAVLFG 297

RESULT 4
PCT-US94-08326-12
Sequence 12, Application PC/TUS9408326
GENERAL INFORMATION:
APPLICANT: North American Vaccine, Inc.
APPLICANT: 12103 Indian Creek Court
APPLICANT: Beltsville, MD 20705
APPLICANT: Pullen, Jeffrey K.
APPLICANT: Soper, Thomas S.
TITLE OF INVENTION: A Method For The High Level
Expression,

```

? NUMBER OF SEQUENCES: 14
? CORRESPONDENCE ADDRESSES:
? ADDRESS: Sterne, Kessler, Goldstein & Fox
? STREET: 1100 New York Avenue, Suite 600
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20005-3934
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/08326
? FILING DATE: Herewith
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/096,181
? FILING DATE: 23-JULY-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Esmond, Robert W.
? REFERENCE/DOCKET NUMBER: 1438.001PC01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 342 amino acids
? TYPE: amino acids
? TOPOLOGY: linear
? MOLECULE TYPE: protein
PCT-US94-08326-14

Query Match          5.9%; Score 79; DB 5; Length 342;
Best Local Similarity 20.1%; Pred. No. 0.83;
Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

QY 28 IETSRHKT-----EDEFQMILRSLP-----DHSGLMFEQIDGIITSSVVPPIIMFL 73
   :: ||| : |::||| : |::||| : |::||| : |::||| : |::||| : |::||| : |
Db 100 VILGRAKTIADGITSAEDEKREYGLNNSDIIPISGNIVGTGTFKIGDVLGA---NYLLAO 156
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 EEMCKRYPHIEQIYGPCKRTGLNTIKYDNPKVEYGADRIVNAVAAILHLYGNPLIVDFGTA 133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 KREGAKGEKKRRNDRAAGEVRIG--EINNGIOVGAKYDANDIAVKIAYGR----- 203
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 TTYCY-VIDENPOVGAATAP-----GITSTEAUYSRAKLPE-----R 170
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 204 TNYKNNEDEHKQOLINGVIATIGTGFSDGLGLVSLDSGIATKTKNKRIKHKEKRYFVSPGFO 263
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 171 IEITRPDNIIG----KNTV----SAMOSGILIF 195
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 264 YELMEDTINVYGNFKYERTSVDOGEXTBRQAVALFG 297
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-08-096-181A-8
? Sequence 8, Application US/08096181A
? Patent No. 6153406
? GENERAL INFORMATION:
? APPLICANT: Tai, Joseph Y.
? APPLICANT: Pullen, Jeffrey K.
? APPLICANT: Soper, Thomas S.
? APPLICANT: Liang, Shu-Mei
? TITLE OF INVENTION: A Method For The High Level Expression,
? TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane Protein
? NUMBER OF INVENTION: P2 From Haemophilus Influenzae Type b
? CORRESPONDENCE ADDRESS:
? ADDRESS: Sterne, Kessler, Goldstein & Fox
? STREET: 1100 New York Avenue, Suite 600
? CITY: Washington
? STATE: D.C.
```

COUNTRY: U.S.A.
 ZIP: 20005-3934
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/096,181A
 FILING DATE: 23-Jul-1993
 CLASSIFICATION: 424
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 S-08-096-181A-8

Query Match 5.9%; Score 79; DB 4; Length 361;
 Best Local Similarity 20.1%; Pred. No. 0.91; Indels 62; Gaps 10;

Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

28 IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIISVVPIMFAL 73
 DB 119 VKLGRAKTIDAGITSADREKYGVLNNSDYIPTSGNTVGTFKIGDVLGA---NYLLAQ 175
 QY 74 ERMCKYFHIEPQIVGPMKGTGLNPKYKVGADRIYNAVAIHLNPLIVDQGT 133
 DB 176 KREGAKGKRNKPDKAGEVRIG---EINNGIYQAKYDADIVAKIAYGR----- 222
 QY 134 TTYCY--IDENKQYMGAIAP-----GITISTEALYSRAALP-----R 170
 DB 223 TNYKYNSEDEHKQOLNGVLAFLGTFPSDGLVLSDSGYAKTKNYKIKHEKRYFVSPGFQ 282
 QY 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
 DB 283 YEIMEDTNYVGNFKYERTSVDOGEKTRQAVLFG 316

RESULT 7

PCT-US94-08326-8

; Sequence 8, Application PC/TUS9408326

; GENERAL INFORMATION:

APPLICANT: North American Vaccine, Inc.

APPLICANT: 12103 Indian Creek Court

APPLICANT: Beltsville, MD 20705

APPLICANT: Pullen, Jeffrey K.

APPLICANT: Liang, Shu-Mei

TITLE OF INVENTION: A Method For The High Level

Purification And Refolding Of The Outer Membrane

TITLE OF INVENTION: Protein

TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/08326

FILING DATE: Herewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/096,181
 FILING DATE: 23-JULY-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REFERENCE/DOCKET NUMBER: 1438.001PC01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-08326-8

Query Match 5.9%; Score 79; DB 5; Length 361;
 Best Local Similarity 20.1%; Pred. No. 0.91; Indels 62; Gaps 10;

Matches 43; Conservative 38; Mismatches 71; Indels 62; Gaps 10;

28 IETSRHKT-----EDEFGMILRSLF-----DHSGLMFEQIDGIISVVPIMFAL 73
 DB 119 VKLGRAKTIDAGITSADREKYGVLNNSDYIPTSGNTVGTFKIGDVLGA---NYLLAQ 175
 QY 74 ERMCKYFHIEPQIVGPMKGTGLNPKYKVGADRIYNAVAIHLNPLIVDQGT 133
 DB 176 KREGAKGKRNKPDKAGEVRIG---EINNGIYQAKYDADIVAKIAYGR----- 222
 QY 134 TTYCY--IDENKQYMGAIAP-----GITISTEALYSRAALP-----R 170
 DB 223 TNYKYNSEDEHKQOLNGVLAFLGTFPSDGLVLSDSGYAKTKNYKIKHEKRYFVSPGFQ 282
 QY 171 IEITRPDNIIG-----KNTV-----SAMQSGILFG 195
 DB 283 YEIMEDTNYVGNFKYERTSVDOGEKTRQAVLFG 316

RESULT 8

US-08-096-181A-10

; Sequence 10, Application US/08096181A

; Patent No. 6153406

; GENERAL INFORMATION:

APPLICANT: Tai, Joseph Y.

APPLICANT: Pullen, Jeffrey K.

APPLICANT: Soper, Thomas S.

TITLE OF INVENTION: A Method For The High Level Expression,

Purification And Refolding Of The Outer Membrane Protein

TITLE OF INVENTION: P2 From Haemophilus Influenzae Type b

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/096,181A

FILING DATE: 23-Jul-1993

CLASSIFICATION: 424

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 10:

LENGTH: 363 amino acids
TYPE: amino acid

US-08-936-135-6

QY 8 GNTNIVLVGYHDDCKLEYHHWRKETSRRHKTEDEFGMLRSLFDHSG---LMFEQID---GI 60
| : | | : | : || : ||| : :||

Db 655 GSHKTECHWEHSDHAOLRNSVLTSK-----TGPIDHTGDNFTTYSADENQKG 704

QY 61 IISVYVPIMFALERMK-TKYFPIPOIYVPGMTGLNKKYDNPREVAGADRIVNAVA--- 116

Db 705 VARLSPVYSSSAHCHMTFWHMGSHNG---TLRVKLRQKPREY--DQVWVYVGHQ 759

QY 117 -----AIHLXGNPLIYVDFGTATTCYIDENKQYMGALPGLTISLEALXS 163

Db 760 GDHWKEGRVILHRSKLXYO-----VIFEG-----EIKGNLGGIAVDISINNHISOE 807

QY 164 RAAKLPRITRPNITKNKT-VSAMOSGILFGYGOVEG 202

Db 808 DCAK-----PTDLKKNTEIKIDETSTGYEGEGEG 839

RESULT 11

US-07-944-943-2

Sequence 2, Application US/07944943

Patent No. 5518724

GENERAL INFORMATION:

APPLICANT: SNYDER, DAVID B.

APPLICANT: VAKHARIA, VIKRAM

TITLE OF INVENTION: NOVEL INFECTIOUS BURSAL DISEASE VIRUS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/944,943

FILING DATE: 19920915

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kelber, Steven B.

REGISTRATION NUMBER: 30,073

REFERENCE/DOCKET NUMBER: 2284-029-0 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1012 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-944-943-2

Query Match 5.8%; Score 78; DB 1; Length 1012;

Best Local Similarity 23.8%; Pred. No. 6.1; Indels 36; Gaps 6;

Matches 39; Conservative 22; Mismatches 67;

QY 97 NIKYDNPREVAGADRIVNAVAIHLXGNPLIYVDF-GTATTTCYIDENKQYMGALPGLT 155

Db 233 NIDAITSLSVGELVFKTSVQSLVGLATILYLGFDGTAVITRAVAANN-----GLT 283

QY 156 ISTEALYSRAAKLPRIETRPDNIIGKNTVSAMOSGILFGYGOVEGVYKRMKQAKODL 215

Db 284 AGTDNLMPENLVIPTNEITOP-----ITSIKIKIVTSKSGGLEG--DQMSWSASGSL 333

QY 216 KYIATGGLAPLIANESDDIDIVDFLLKGLLEI-YERNRVGSV 258

Db 334 AVTIHGNNP-----GALRPVTLVAIVERATGVS 362

RESULT 12

US-08-219-262B-2

Sequence 2, Application US/08219262B

Patent No. 5788970

GENERAL INFORMATION:

APPLICANT: VAKHARIA, VIKRAM

APPLICANT: SNYDER, DAVID B

APPLICANT: MENDEL-WHERSAT, STEPHANIE A

TITLE OF INVENTION: CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS

TITLE OF INVENTION: CDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED THEREON

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/219,262B

FILING DATE: 29-MAR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 2747-047-27

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1012 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Infectious bursal disease virus

STRAIN: DS326

US-08-219-262B-2

Query Match 5.8%; Score 78; DB 1; Length 1012;

Best Local Similarity 23.8%; Pred. No. 6.1; Indels 36; Gaps 6;

Matches 39; Conservative 22; Mismatches 67;

QY 97 NIKYDNPREVAGADRIVNAVAIHLXGNPLIYVDF-GTATTTCYIDENKQYMGALPGLT 155

Db 233 NIDAITSLSVGELVFKTSVQSLVGLATILYLGFDGTAVITRAVAANN-----GLT 283

QY 156 ISTEALYSRAAKLPRIETRPDNIIGKNTVSAMOSGILFGYGOVEGVYKRMKQAKODL 215

Db 284 AGTDNLMPENLVIPTNEITOP-----ITSIKIKIVTSKSGGLEG--DQMSWSASGSL 333

QY 216 KYIATGGLAPLIANESDDIDIVDFLLKGLLEI-YERNRVGSV 258

Db 334 AVTIHGNNP-----GALRPVTLVAIVERATGVS 362

RESULT 13

US-09-031-655-2

Sequence 2, Application US/09031655

Patent No. 6017759

GENERAL INFORMATION:

APPLICANT: VAKHARIA, VIKRAM

APPLICANT: SNYDER, DAVID B

```

      ; EARLIER APPLICATION NUMBER: 08/693,247
      ; EARLIER FILING DATE: 1996-08-16
      ; NUMBER OF SEQ ID NOS: 36
      ; SOFTWARE: PatentIn Ver. 2.0
      ; SEQ ID NO 2
      ; LENGTH: 1213
      ; TYPE: PRT
      ; ORGANISM: Classical Swine Fever Virus
US-08-937-102-2

Query Match          5.8%; Score 78; DB 2; Length 1213;
Best Local Similarity 20.7%; Pred. No. 8.1;
Matches 45; Conservative 35; Mismatches 73; Indels 64; Gaps 11

QY  20  CKLEHW-RIETSRKTEDEFEQMILR-----SLEDSGLMFEQID--GIITSSVYP 67
      |::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB  993  GSKDHYDLQAKRGIEIDGINITKSFREMYNDWSLYEDSLMITOLELNNLLISEEP 1042
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY  68  PIMFLEBRCYKFFHIEP-----OIVGPMKTC-LNIVYDNPKKEGADRIYVA 114
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB  1043 ---MAVKNIMARTDHPEPIQLAVNSYETQVPLFKIRNGEVTDTDYDNTFFLNARKLGD 1099
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY  115  VAATHLYGNPLIVDFGATTTCYIDENKQYVGAIA----PGITISTEALYSRAAKLP 169
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB  1100  V-----PPVYATEDEDELAVELLGLDMDPDPGNGTVEA--GRALK-- 1137
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY  170  RLEITRPDNIIGKNYVSAMOSGILEGYYGQVGEIYKR 206
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB  1138 -----QVVGSLTAENALLVALFEGYVG--YQALSKR 1165
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 15
US-08-876-991-2
Sequence 2, Application US/08876991
Patent No. 5825360
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESSEE: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,991
FILING DATE: 16-JUN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/747,577
FILING DATE:
APPLICATION NUMBER: US/08/650,584
FILING DATE:
APPLICATION NUMBER: US/08/469,702
FILING DATE:
APPLICATION NUMBER: US/08/123,596
FILING DATE:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone

REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3898 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-876-991-2

Query Match 5.8%; Score 78; DB 2; Length 3898;
Best Local Similarity 20.7%; Pred. No. 51;
Matches 45; Conservative 35; Mismatches 73; Indels 64; Gaps 11;

QY 20 GLEIHW-RIETSRKTEDEFGMLR-----SLFDHSGIMFEQID--GIITSSVPE 67
DB 2113 GSKDYHYDLQAOYRGIEDGINITSFRBMNTDWSLYEEDSLMITOLETLNNLLISELP 2172
Y 68 PIMFALERMCTKYFHIEP-----QIVPGMKTG-LNIKYDNPKVEGADRIVNA 114
DB 2173 ---MAVKNIMARDHEPIQLAVNSYETQVPVLPKIRNGEYTDFTYDNTFLNARKLGDD 2229
QY 115 VAAIHLYGNPLIVDFGTRATTCYIDENKQYMGATA-----PGITISTEALXSPRAKLP 169
DB 2230 V-----PPVYATEDEDLAVELLGLDMPDPGNOGTVEA--GRALK-- 2267
QY 170 RIETRPDNIIGKNIVSAMOSGILFGYGVGEIVKR 206
DB 2268 -----QVVGSLTAENALVALFQYVG-YQALSKR 2295

Search completed: April 21, 2003, 13:43:34
Job time : 33 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2003, 13:40:06 ; Search time 17 Seconds

(Without alignments)
1147.552 Million cell updates/sec

Title: US-09-813-453A-2

Perfect score: 1335
Sequence: 1 LLLVIVGNTNVLGYHDK.....PFLTLKGLLEIYERNVGSV 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 288829 segs, 75613885 residues

Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1335	100.0	258	9	US-09-813-453A-2
2	1114	83.4	233	9	US-09-813-453A-17
3	1046	78.4	258	9	US-09-813-453A-49
4	1034	77.5	262	9	US-09-813-453A-45
5	987	73.9	254	9	US-09-813-453A-47
6	794.5	59.5	256	9	US-09-813-453A-55
7	756	56.6	255	9	US-09-813-453A-7
8	655	49.1	265	9	US-09-813-453A-4
9	646.5	48.4	250	9	US-09-813-453A-3
10	577	43.2	260	9	US-09-813-453A-51
11	537	40.2	258	9	US-09-813-453A-56
12	515	38.6	219	9	US-09-813-453A-57
13	494.5	37.0	272	9	US-09-813-453A-5
14	493.5	37.0	272	9	US-09-813-453A-5
15	428	32.1	262	9	US-09-812-363-276
16	405.5	30.4	246	9	US-09-813-453A-8
17	338.5	25.4	212	9	US-09-813-453A-9
18	327.5	24.5	273	9	US-09-813-453A-59
19	319	23.9	257	9	US-09-813-453A-10

20	271.5	20.3	262	9	US-09-813-453A-11	Sequence 11, App1
21	207	15.5	244	9	US-09-813-453A-41	Sequence 41, App1
22	203	15.2	241	9	US-09-813-453A-63	Sequence 63, App1
23	163	12.2	249	9	US-09-813-453A-70	Sequence 70, App1
24	163	12.2	257	9	US-09-813-453A-13	Sequence 13, App1
25	154.5	11.6	229	9	US-09-813-453A-12	Sequence 12, App1
26	154	11.5	249	9	US-09-813-453A-61	Sequence 61, App1
27	151	11.3	460	9	US-09-813-453A-39	Sequence 39, App1
28	150	11.2	592	9	US-09-813-453A-22	Sequence 22, App1
29	150	11.2	592	9	US-09-813-453A-65	Sequence 65, App1
30	134.5	10.1	242	9	US-09-813-453A-14	Sequence 14, App1
31	133	10.0	248	9	US-09-813-453A-20	Sequence 20, App1
32	129.5	9.7	267	9	US-09-813-453A-15	Sequence 15, App1
33	109.5	8.2	223	9	US-09-813-453A-74	Sequence 74, App1
34	109.5	8.2	223	9	US-09-813-453A-14	Sequence 14, App1
35	109.5	8.2	223	9	US-09-813-453A-67	Sequence 67, App1
36	109	8.2	209	9	US-09-813-453A-21	Sequence 21, App1
37	80	6.0	449	10	US-09-815-242-5474	Sequence 5474, App
38	80	6.0	449	10	US-09-815-242-12348	Sequence 12348, A
39	80	6.0	449	10	US-09-815-242-12793	Sequence 12793, A
40	79	5.9	636	10	US-09-205-658-160	Sequence 160, App
41	78.5	5.9	337	9	US-10-075-846-12	Sequence 12, App1
42	78.5	5.9	337	9	US-09-815-242-13811	Sequence 13811, A
43	78.5	5.9	1057	10	US-09-815-242-5798	Sequence 5798, App
44	78.5	5.9	1107	10	US-09-815-242-12815	Sequence 12815, A
45	78.5	5.9	1198	10	US-09-815-242-12446	Sequence 12446, A

ALIGNMENTS

RESULT 1
US-09-813-453A-2
Sequence 2, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: ANTIBIOTICS
FILE REFERENCE: 062-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 258
TYPE: PRT
ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 100.0% ; Score 1335 ; DB 9 ; Length 258 ;
Best Local Similarity 100.0% ; Pred. No. 5.4e-130 ;
Matches 258 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;

QY 1 LLLVIVGNTNVLGYHDKLEIYHRIETSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
DB 1 LLLVIVGNTNVLGYHDKLEIYHRIETSRHKTEDEFGMLRSLFDHSGLMFEQIDGI 60
QY 61 IISVVPPIIMFALERMCTYFPIIEPOIVPGKKTGINIKYDNPKEVGADRIYNAVAIHL 120
DB 61 IISVVPPIIMFALERMCTYFPIIEPOIVPGKKTGINIKYDNPKEVGADRIYNAVAIHL 120
QY 121 YGNPILVDFGATYTCYIDENKQYMGALAGITISTEALYSRAKLPRIETIPPDNI 180
DB 121 YGNPILVDFGATYTCYIDENKQYMGALAGITISTEALYSRAKLPRIETIPPDNI 180
QY 181 GNTVSAMSGILFVGQVEGIVRBMKQAKODAKVATGTGLAPLINESDCIDIYDPF 240
DB 181 GNTVSAMSGILFVGQVEGIVRBMKQAKODAKVATGTGLAPLINESDCIDIYDPF 240

Db 181 GKNVSAWOSGILFGYGVGEGIVKRMKQAKODLKVATGGLAPLIANESDCIDIYDPF 240

OY 241 LTLKGLLEIYERNRGSV 258
Db 241 LTLKGLLEIYERNRGSV 258

RESULT 2
US-09-813-453a-17
Sequence 17, Application US/09813453A

Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OCG-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 233
TYPE: PRF
ORGANISM: Bacillus subtilis
US-09-813-453a-17

Query Match 83.4%; Score 1114; DB 9; Length 233;
Best Local Similarity 99.5%; Pred. No. 3.2e-107;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LLLVLDGNTNTVIGYVHDSKLEFHWRIETSRHKTDEDFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLLVLDGNTNTVIGYVHDSKLEFHWRIETSRHKTDEDFGMLRSLFDHSGLMFEQIDGI 60
OY 61 IISVVPIMFALERMCTKTFHIEPQIVGPMKTGMLIKNDKREKAGADRIYNAVAIHL 120
Db 61 IISVVPIMFALERMCTKTFHIEPQIVGPMKTGMLIKNDKREKAGADRIYNAVAIHL 120
OY 121 YGNPLIVDFGATTCYCIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
Db 121 YGNPLIVDFGATTCYCIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
OY 181 GKNVSAWOSGILFGYGVGEGIVKRMKQAKOD 214
Db 181 GKNVSAWOSGILFGYGVGEGIVKRMKQAKOD 214

RESULT 3
US-09-813-453a-49
Sequence 49, Application US/09813453A

Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OCG-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 49
LENGTH: 258
TYPE: PRF

ORGANISM: Bacillus steatothermophilus
US-09-813-453a-49

Query Match 78.4%; Score 1046; DB 9; Length 258;
Best Local Similarity 78.3%; Pred. No. 3.9e-100;
Matches 198; Conservative 32; Mismatches 23; Indels 0; Gaps 0;

OY 1 LLLVLDGNTNTVIGYVHDSKLEFHWRIETSRHKTDEDFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLLVLDGNTNTVIGYVHDSKLEFHWRIETSRHKTDEDFGMLRSLFDHSGLMFEQIDGI 60
OY 61 IISVVPIMFALERMCTKTFHIEPQIVGPMKTGMLIKNDKREKAGADRIYNAVAIHL 120
Db 61 IISVVPIMFALERMCTKTFHIEPQIVGPMKTGMLIKNDKREKAGADRIYNAVAIHL 120
OY 121 YGNPLIVDFGATTCYCIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
Db 121 YGNPLIVDFGATTCYCIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
OY 181 GKNVSAWOSGILFGYGVGEGIVKRMKQAKODLKVATGGLAPLIANESDCIDIYDPF 240
Db 181 GKNVSAWOSGILFGYGVGEGIVKRMKQAKODLKVATGGLAPLIANESDCIDIYDPF 240
OY 241 LTLKGLLEIYERN 253
Db 241 LTLKGLLEIYERN 253

RESULT 4
US-09-813-453a-45
Sequence 45, Application US/09813453A

Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: OCG-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 45
LENGTH: 262
TYPE: PRF
ORGANISM: Bacillus anthracis
US-09-813-453a-45

Query Match 77.5%; Score 1034; DB 9; Length 262;
Best Local Similarity 75.1%; Pred. No. 7e-99; 29; Indels 0; Gaps 0;
Matches 190; Conservative 34; Mismatches 29; Indels 0; Gaps 0;

OY 1 LLLVLDGNTNTVIGYVHDSKLEFHWRIETSRHKTDEDFGMLRSLFDHSGLMFEQIDGI 60
Db 1 MLLVLDGNTNTVIGYVHDSKLEFHWRIETSRHKTDEDFGMLRSLFDHSGLMFEQIDGI 60
OY 61 IISVVPIMFALERMCTKTFHIEPQIVGPMKTGMLIKNDKREKAGADRIYNAVAIHL 120
Db 61 IISVVPIMFALERMCTKTFHIEPQIVGPMKTGMLIKNDKREKAGADRIYNAVAIHL 120
OY 121 YGNPLIVDFGATTCYCIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
Db 121 YGNPLIVDFGATTCYCIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
OY 181 GKNVSAWOSGILFGYGVGEGIVKRMKQAKODLKVATGGLAPLIANESDCIDIYDPF 240
Db 181 GKNVSAWOSGILFGYGVGEGIVKRMKQAKODLKVATGGLAPLIANESDCIDIYDPF 240
OY 241 LTLKGLLEIYERN 253
Db 241 LTLKGLLEIYERN 253

Db 241 LTLKGLMYERN 253

RESULT 5
US-09-813-453A-47

Sequence 47, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: CGZ-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 47
LENGTH: 254
TYPE: PRT
ORGANISM: Bacillus halodurans
US-09-813-453A-47

Query Match 73.9%; Score 987; DB 9; Length 254;
Best Local Similarity 72.3%; Pred. No. 4.8e-94;
Matches 183; Conservative 33; Mismatches 37; Indels 0; Gaps 0;

QY 1 LLLVIVDGNNTVGVYHDGKLEYHWRHETSRHKTDEFGMLNSLPDHSGLMFEQIDGI 60
Db 1 MLVIVDGNNTVGVYHDEFLVHWRHETSRHKTDEFGMLNSLPDHSGLMFEQIDGI 60
QY 61 IISVPPIMFALERMCTKRYHIEPOIVPGMKTGLNRYDNPKEVGADRIYNAVAATHL 120
Db 61 VISSVPPIMFALERMCTKRYHIEPOIVPGMKTGLNRYDNPKEVGADRIYNAVAATHL 120
QY 121 YGNPLIVDFGTATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
Db 121 YGPAIVDFGTATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
QY 181 GKNTVSAMQSGILFGYGOVEGIVRKMKQAKODLVATGGLAPLANESDCIDIYDPF 240
Db 181 GTNTIDMSQSGILFGYGOVEGIVRKMKQAKODLVATGGLAPLANESDCIDIYDPF 240
QY 241 LTLKGLMYERN 253
Db 241 LTLKGLMYERN 253

RESULT 6
US-09-813-453A-55
Sequence 55, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: CGZ-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 55
LENGTH: 256

TYPE: PRT
ORGANISM: Clostridium difficile
US-09-813-453A-55

Query Match 59.5%; Score 794.5; DB 9; Length 256;
Best Local Similarity 60.8%; Pred. No. 3.7e-74;
Matches 155; Conservative 40; Mismatches 59; Indels 1; Gaps 1;

QY 1 LLLVIVDGNNTVGVYHDGKLEYHWRHETSRHKTDEFGMLNSLPDHSGLMFEQIDGI 60
Db 1 MLVIVDGNNTVGVYHDEFLVHWRHETSRHKTDEFGMLNSLPDHSGLMFEQIDGI 60
QY 61 IISVPPIMFALERMCTKRYHIEPOIVPGMKTGLNRYDNPKEVGADRIYNAVAATHL 120
Db 61 VISSVPPIMFALERMCTKRYHIEPOIVPGMKTGLNRYDNPKEVGADRIYNAVAATHL 120
QY 121 YGNPLIVDFGTATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
Db 121 YGPAIVDFGTATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
QY 181 GKNTVSAMQSGILFGYGOVEGIVRKMKQAKODLVATGGLAPLANESDCIDIYDPF 239
Db 181 CKSTVSAMQSGILFGYGOVEGIVRKMKQAKODLVATGGLAPLANESDCIDIYDPF 240
QY 240 LTLKGLMYERN 254
Db 241 LTLKGLMYERN 255

RESULT 7
US-09-813-453A-7

Sequence 7, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
FILE REFERENCE: CGZ-001
CURRENT APPLICATION NUMBER: US/09/813,453A
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 255
TYPE: PRT
ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 56.6%; Score 756; DB 9; Length 255;
Best Local Similarity 57.3%; Pred. No. 3.3e-70;
Matches 146; Conservative 46; Mismatches 62; Indels 0; Gaps 0;

QY 1 LLLVIVDGNNTVGVYHDGKLEYHWRHETSRHKTDEFGMLNSLPDHSGLMFEQIDGI 60
Db 1 MLVIVDGNNTVGVYHDEFLVHWRHETSRHKTDEFGMLNSLPDHSGLMFEQIDGI 60
QY 61 IISVPPIMFALERMCTKRYHIEPOIVPGMKTGLNRYDNPKEVGADRIYNAVAATHL 120
Db 61 VISSVPPIMFALERMCTKRYHIEPOIVPGMKTGLNRYDNPKEVGADRIYNAVAATHL 120
QY 121 YGNPLIVDFGTATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
Db 121 YGPAIVDFGTATTCYIDENKQYMGALAPGITISTEALYSRAKLPRIETRPDNI 180
QY 181 GKNTVSAMQSGILFGYGOVEGIVRKMKQAKODLVATGGLAPLANESDCIDIYDPF 240
Db 181 ARNTVSAMQSGILFGYGOVEGIVRKMKQAKODLVATGGLAPLANESDCIDIYDPF 240

```
QY      241  LTLEGLIYERNR 254
          |||:| :| |||
Db      241  LTLEGLRIYERNR 254
```

RESULT 8
US-09-813-453A-4

Patent No. US20020168681A1

GENERAL INFORMATION:

APPLICANT: Yocum, R. Rogers

APPLICANT: Patterson, Thomas A.

TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF

; TITLE OF INVENTION: ANTIBIOTICS

FILE REFERENCE: OGG-001

; CURRENT APPLICATION NUMBER: US/09/813,453A

; CURRENT FILING DATE: 2001-03-20

PRIOR APPLICATION NUMBER: US 60/227,860

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: 09/667,569

PRIOR FILING DATE: 2000-09-21

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn Ver. 2.0

SEO ID NO 4

LENGTH: 265

TYPE: PRT

ORGANISM: *Streptomyces coelicolor*

ORGANISM: SL
HS-09-813-453A-4

Query Match	49.18	Score 655	DB 9	length 265
Best Local Similarly	51.08	Pred. NO. 1e-59		
Matches 133	Conservative 46	Mismatches 74	Indels 8	Gaps 3

Qy	1	LLVTDVCNTFTVLGVYHDXKLEHYHMRIETSRHKTEDEFGMLIRSLFDHSGLMEEQ----	56
Dd	1	MLLTIDGENTFTVLGLFEGEDIVBHRMSTDSRRTABELAVLDDLGCMHPLDDELGDG	60
Qy	57	IDGIILSSVPPIMFATERMCTKYFHIEPQI-VGPCKMTGLINIKYDNPKREVGADRIYNAV	115
Dd	61	IDGIALCATVSYLHELREVTRRYYGDVPALVLPBGWKGTVPLITDHPKEVGADRINNAV	120
Qy	116	AAILHYGNRPFLVNDPGFRATTCYCIDENKQYNGMALAGTISTEALYSRAAKLPRIETR	175
Dd	121	AAVELYGGPALVNDPGFRATTITDAASAREYIGSVIAPGISLVALGVKGQOLKEITVAR	180
Qy	176	PDNIIKKNTVASAMSGILFIYGVOGEVIGKRMMQA--KODLVIAITGGLAPIANESD	233
Dd	181	PRSYVIGNKTVEAMSGIYIFGAQVDGVNNMARELADDPDVVIAITGTGLAPVYLESS	240
Qy	233	CIDIYDPRLTLKGLELIYEARN	233
Dd	241	VIDEHEPWLITLMGLRLVYEARN	261

```

: RESULT 9
: US-09-813-453A-3
: Sequence 3, Application US/09813453A
: Patent No. US20020168681A1
: GENERAL INFORMATION:
: APPLICANT: Yocum, R. Rogers
: APPLICANT: Patterson, Thomas A.
: TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
: TITLE OF INVENTION: ANTIBIOTICS
: FILE REFERENCE: OC2-001
: CURRENT APPLICATION NUMBER: US/09/813,453A
: CURRENT FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: US 60/227,860
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: 09/667,569
: PRIOR FILING DATE: 2000-09-21
: NUMBER OF SEQ ID NOS: 77
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 3

```

```

; LENGTH: 250
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

```

Query Match	48.4%	Score 646.5;	DB 9;	Length 250;
Best Local Similarity	53.9%	Pred. No. 7.1e-59;		
Matches 125; Conservative	46;	Mismatches 56;		Gaps 1

[illegible]

```

RESULT 10
US-09-813-453A-51
: Sequence 51, Application US/09813453A
: Patent No. US20020168681A1
: GENERAL INFORMATION:
: APPLICANT: Yocum, R. Rogers
: APPLICANT: Paterson, Thomas A.
: TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
: TITLE OF INVENTION: ANTIBIOTICS
: FILE REFERENCE: CG2-001
: CURRENT APPLICATION NUMBER: US/09/813,453A
: CURRENT FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: US 60/227,860
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: 09/667,569
: PRIOR FILING DATE: 2000-09-21
: NUMBER OF SEQ ID NOS: 77
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 51
: LENGTH: 260
: TYPE: FRT
: ORGANISM: Caulobacter crescentus
US-09-813-453A-51

```

	Query Match	43.28;	Score 577;	DB 9;	Length 260;	
	Best Local Similarity	44.4%;	Pred. No. 1.1e-51;			
	Matches 114;	Conservative	50;	Mismatches 91;	Indels 2;	Gaps 1
Oy	1 LLLVIVDVGNTNVLGVYHDKLELHYHRIRIETSHHKTEDEFGMLRLSLFDHSGLFMFEDIGI 60	:	:	:	:	:
Dd	1 MLLALEGGTNTNMFAIHDSASWAOQRSATESTRADDEVVMLSQLSSMGJGFRADAV 60	:	:	:	:	:
Oy	61 IISVVPPIMFALERMCKTFPHIEPOIVGPGMKTGILNKIKINVDNEKEVGADRIVNAVAAIL 120	:	:	:	:	:
Dd	61 IISSVVPQSIIFNLRLNSRRFYFNVEPLVIGENAKLGIDVRLEKRSSEAGADRLVANIGAAMV 120	:	:	:	:	:
Oy	121 YGNPLIVDFGAITYCYIDENKQYWGAIAEGITISTALYSRAKLRIETTRP--DN 178	:	:	:	:	:
Dd	121 YPGPLVIDISGATTFEDIAAADGAEEGFIAPGINLSLMHLAAKLPRIAIOBPANR 180	:	:	:	:	:
Oy	179 IIGKTVSAMOGSILFEGYGVEGYEITKRKKMQAKODLKVIATNGCLAPLANESDCIDIYD 238	:	:	:	:	:
Dd	181 IYGTIVTSAMOSGVFWGYISLLEGIVARIKAERGEPFVIATGVSASFLEGATSDSIDHD 240	:	:	:	:	:
Oy	239 PELLTKGLELIERNRV 255	:	:	:	:	:
Dd	241 SDLTIIRGLLEIYRRNTI 257	:	:	:	:	:

RESULT 11

US-09-813-453A-6
Sequence 6, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: ANTIBIOTICS
FILE REFERENCE: OGT-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 258
TYPE: PRN
ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

Query Match 40.2%; Score 537; DB 9; Length 258;
Best Local Similarity 43.9%; Pred. No. 1,5e-47;
Matches 112; Conservative 46; Mismatches 95; Indels 2; Gaps 2;

OY 1 LLLVIVDGNNTNTVLCVYHDGKLEYHWRITSRHKTDEDEFGMLSLPDSGLMFEQIDGI 60
DB 1 MLLCIDCGTNTVSVWCGTDFATMTATDHRFADYFVWLTMLQKCLQ-GRISEA 59
OY 61 IISVVPPIEMALERMCTKYFHEIPOLV-PGMKTGLNIXKDNKKEVGADRIYVNAVAIH 119
DB 60 IISSTAPRVENFLVLCGRYFDCRPYVGRKPCLPVAPRPDPTVGPDRLVNTVAGYD 119
OY 120 LYGNPLVDFGTAATYCYIDENKQYMGALAPGITTSTALYSRAKLPRIETRPDNI 179
DB 120 RHGGDLVYDGTATPTDVPAPDAYIGGVYAPGVNLSLEALHMAAALPHVDYTKRGV 179
OY 180 IGMTVSAMQSGILFGYGVGVEGIYVRKMKQAKODLVYATGGLAPLIANESCDIYDP 239
DB 180 IGMTVACIOGSGVYWGVIQLEGIYRQIRMERDRPKVITATGLASLFDLGFDEKVED 239
OY 240 FLTKGLELYERNR 254
DB 240 DLTMHGRLPLFDYNK 254

RESULT 12

US-09-813-453A-57
Sequence 57, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: ANTIBIOTICS
FILE REFERENCE: OGT-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 57
LENGTH: 219
TYPE: PRN
ORGANISM: Dehalococcoides ethenogenes

US-09-813-453A-57

Query Match 38.6%; Score 515; DB 9; Length 219;
Best Local Similarity 45.8%; Pred. No. 2.3e-45;
Matches 97; Conservative 45; Mismatches 70; Indels 0; Gaps 0;

OY 2 LLLVIVDGNNTNTVLCVYHDGKLEYHWRITSRHKTDEDEFGMLSLPDSGLMFEQIDGI 61
DB 5 LVAVDIGNTSVNICIFEECKLAWHLSVAQRADYASILLGLLOHAGIHEELNRVI 64
OY 62 IISVVPPIEMALERMCTKYFHEIPOLV-PGMKTGLNIXKDNKKEVGADRIYVNAVAIH 121
DB 65 MCSVVPPLTTTFEEVFKSYFAAPLVYAGIKSGVYKVRMNPREGVADRIVNAARVL 124
OY 122 GNPPLVDFGTAATYCYIDENKQYMGALAPGITTSTALYSRAKLPRIETRPDNI 181
DB 125 PGACIIVDMGTATPTDLSRGAYIGALAPGATSAQAIAEKTSKLPKTEIRPAKVI 184
OY 182 KNTVSAMQSGILFGYGVGVEGIYVRKMKQAKQ 213
DB 185 SNTVSAMQSGIYFGYIGLVEELVRIOTELQ 216

RESULT 13

US-09-813-453A-5
Sequence 5, Application US/09813453A
Patent No. US20020168681A1
GENERAL INFORMATION:
APPLICANT: Yocum, R. Rogers
APPLICANT: Patterson, Thomas A.
TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: ANTIBIOTICS
FILE REFERENCE: OGT-001
CURRENT APPLICATION NUMBER: US/09/813,453A
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: 09/667,569
PRIOR FILING DATE: 2000-09-21
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 272
TYPE: PRN
ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

Query Match 37.0%; Score 494.5; DB 9; Length 272;
Best Local Similarity 38.8%; Pred. No. 4e-43;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

OY 1 LLLVIVDGNNTNTVLCVYHDGKLEYHWRITSRHKTDEDEFGMLSLPDSGLMFEQ 56
DB 1 MLLAIDVRNTHTVGLSGMKEHAKVVOOWRIRTESEVTADELATLIDGLIGEDS--ER 57
OY 57 IDGIISVVPPIEMALERMCTKYFHEIPOLV-PGMKTGLNIXKDNKKEVGADRIYVNAV 115
DB 58 LTGTAALSTVPSVLEHVRIMLDQWPSVPHLIEPVKRTGIPLVDPNPKVEGARIVNCL 117
OY 116 AAHLHXGNPLVDFGTAATYCYIDENKQYMGALAPGITTSTALYSRAKLPRIETRP 175
DB 118 AAYDRFRMAALVDFGSSICVDVYSAKGEFLGAIAGVOVSSDAANAARSALRVELAR 177
OY 176 PNTIIGKNTVSAMQSGILFGYGVGVEGIYVRM-----QAKODLKYATGGLAPLIANE 230
DB 178 PRSVYKNTVECMQAGAVFAGLVDELVGRIREDVSGFVSDHVAIVATGHTAPLLPE 237
OY 231 SDCIDIYDPELTTLKGLIELYERN 253
DB 238 LHTVDHYDQHLTLQGLRLVEERN 260

RESULT 14

[illegible][illegible]